

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 17, 2003, 19:37:48 ; Search time 2944 Seconds

(without alignments)
10676.301 Million cell updates/sec

Title: US-09-972-758a-1

Perfect score: 1080

Sequence: 1 atgcgcgacgcatctctgtc.....tttccaagtgttgagactag 1080

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_tod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1076.8	99.7	3624	9	AB021179	AB021179 Homo sapi
2	1076.8	99.7	100836	2	AC087298	AC087298 Homo sapi
3	1075.2	99.6	2178	9	AK023624	AK023624 Homo sapi
4	1073.6	99.4	2086	9	BC006460	BC006460 Homo sapi
5	1052.8	97.5	142326	2	AC024047	AC024047 Homo sapi
6	840.8	77.9	3402	10	AY090614	AY090614 Mus muscu
7	840.8	77.9	111002	10	AL731805	AL731805 Mouse DNA
8	833.8	77.2	1488	10	BC022111	BC022111 Mus muscu
9	793.8	73.5	175120	2	AC120950	AC120950 Rattus no
10	355.2	32.9	461	6	AX331371	AX331371 Sequence
11	174.8	16.2	1238	9	BC025970	BC025970 Homo sapi
12	174.8	16.2	1330	9	AK056946	AK056946 Homo sapi
13	174.8	16.2	116035	2	AC008105	AC008105 Homo sapi
14	173.2	16.0	2048	9	AK096785	AK096785 Homo sapi
15	172	15.9	223442	10	AL662804	AL662804 Mouse DNA
16	103	9.5	167636	2	AC107097	AC107097 Rattus no
17	99.4	9.2	82400	10	AC090495	AC090495 Genomic s
18	99.2	9.2	195673	2	AL807807	AL807807 Mus muscu
19	87.4	8.1	198228	8	AC102097	AC102097 Mus muscu
20	87	8.1	207683	2	AC098712	AC098712 Mus muscu
21	86.2	8.0	283816	2	AC124587	AC124587 Mus muscu
22	85.8	7.9	194334	2	AC101527	AC101527 Mus muscu
23	84.8	7.9	248037	2	AC121887	AC121887 Mus muscu
24	84.4	7.8	195690	10	AC098726	AC098726 Mus muscu
25	84.2	7.8	155662	2	AC112388	AC112388 Rattus no
26	84	7.8	180418	2	AC106674	AC106674 Rattus no
27	83.4	7.7	231853	2	AC093352	AC093352 Mus muscu
28	83	7.7	225045	2	AL833803	AL833803 Mus muscu
29	82.8	7.7	318930	2	AC073495	AC073495 Mus muscu
30	80.4	7.4	168210	2	AC116700	AC116700 Mus muscu
31	79.6	7.4	184865	2	AL807824	AL807824 Mus muscu
32	79.2	7.3	272545	2	AC090533	AC090533 Mus muscu
33	78.6	7.3	165197	2	AC112538	AC112538 Rattus no
34	78.2	7.2	203982	9	AC067854	AC067854 Homo sapi
35	77.8	7.2	234817	10	AL663048	AL663048 Mouse DNA
36	77.6	7.2	185460	2	AC115552	AC115552 Rattus no
37	77.4	7.2	180850	2	AC107828	AC107828 Mus muscu
38	75.4	7.0	287323	2	AC093363	AC093363 Mus muscu
39	75.2	7.0	97348	10	AF091216	AF091216 Mus muscu
40	74.4	6.9	256346	2	AC105970	AC105970 Mus muscu
41	73.8	6.8	126357	2	AC121480	AC121480 Rattus no
42	73.8	6.8	201271	2	AC124194	AC124194 Mus muscu
43	73.8	6.8	206497	10	AC084389	AC084389 Mus muscu
44	73.2	6.8	186400	2	AC095843	AC095843 Rattus no
45	73.2	6.8	227386	2	AC125235	AC125235 Mus muscu

ALIGNMENTS

RESULT 1
AB021179 3624 bp mRNA linear PRI 05-DEC-2000
LOCUS Homo sapiens mRNA for HEXIM1 protein, complete cds.
DEFINITION
AB021179
ACCESSION
AB021179.1 GI:4062855
VERSION
HEXIM1; HEXIM1 protein.
KEYWORDS
Homo sapiens coronary artery smooth muscle cell cDNA to mRNA.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (sites)
Kusuhara,M., Nagasaki,K., Kimura,K., Maass,N., Manabe,T.,
Ishikawa,S., Aikawa,M., Miyazaki,K. and Yamaguchi,K.
TITLE
Cloning of hexamethylene-bis-acetamide-inducible transcript,

Pred. No. is the number of results predicted by chance to have a

JOURNAL Biomed. Res. 20, 273-279 (1999)
 REFERENCE 2 (bases 1 to 3624)
 AUTHORS Kusunhara, M.
 TITLE Direct Submission
 JOURNAL Submitted (14-DEC-1998) Masatoshi Kusunhara, National Defense
 Medical College, The First Department of Internal Medicine, 3-2
 Namiki, Tokorozawa, Saitama 359-8513, Japan
 (E-mail: mkusu@med.nidmc.ac.jp, Tel: +81-42-995-1597,
 Fax: +81-42-996-5200)

FEATURES
 source
 1. 3624
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="coronary artery smooth muscle cell"

gene
 1. 3624
 /gene="HEXIM1"
 690. 1769
 /gene="HEXIM1"
 /codon_start=1
 /product="HEXIM1 protein"
 /protein_id="BAA36166.1"
 /db_xref="GI:4062856"

/translation="MAPRUSLYOHQROTSMCTGAANYQVELNRRPRPAAERPEED
 SRMSRAFPOLGRRPPEBESLSQPPDTQACPESSCLREBKGMDSSAGD
 FPPAEVPEPTAEALLQPCHEDSASKIAPAAAGEEWAQOQOLQKKRRRPSK
 KRHWKPYKLTWEKKKFKDEKQSLRASKIREMFAKGPVAPYNTQFLMDHDEEP
 DLKGLYSKRAAKSDTSDDDFMEEGEEDGSDGMGDSEFLQDRFSEYRYHT
 ESLQMSKQELIKELYELEKELCSHMEEDNNRLRLSKRLGGDDARVRELELDRLRA
 ENQLTLENELHROERAPLSKFGD"

BASE COUNT 978 a 775 c 964 g 907 t
 ORIGIN

Query Match 99.7%; Score 1076.8; DB 9; Length 3624;
 Best Local Similarity 99.8%; Pred. No. 3.3e-169;
 Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ATGCCGAGCCATTCTTTCAGAAATATACACACACCTCAAACTAGCACTGTACAGT 60
 DB 690 ATGCCGAGCCATTCTTTCAGAAATATACACACACCTCAAACTAGCACTGTACAGT 749
 OY 61 GCTGCTGTCACAGAGAGACTGTACACCTGAGGCGCCCGAGGCGGAGAGGAGGTG 120
 DB 750 GCTGCTGTCACAGAGAGACTGTACACCTGAGGCGCCCGAGGCGGAGAGGAGGTG 809
 OY 121 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 DB 810 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 869
 OY 181 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 DB 870 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929
 OY 241 GAATCTAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 DB 930 GAATCTAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 989
 OY 301 GCGGAGCTCCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 DB 990 GCGGAGCTCCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
 OY 361 CCTGTATGACTCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 DB 1050 CCTGTATGACTCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1109
 OY 421 TGGGAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 480
 DB 1110 TGGGAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1169
 OY 481 AAGCGGCAATGGAAACCGTACTACAGCTGAGTGGGAAGAGAGAGAGAGAGAGAGAGAG 540
 DB 1170 AAGCGGCAATGGAAACCGTACTACAGCTGAGTGGGAAGAGAGAGAGAGAGAGAGAGAG 1229

OY 541 AAACAGAGCCCTTCAGAGCTTCAAGATCCGAGCCAGAGATGTTCCGCAAGGGCCAGCCGTC 600
 DB 1230 AAACAGAGCCCTTCAGAGCTTCAAGATCCGAGCCAGAGATGTTCCGCAAGGGCCAGCCGTC 1289
 OY 601 GCGCCCTTATACACACAGCAGCTTCTCATGATGATACAGCAGCAGAGAGAGAGCCGATCTC 660
 DB 1290 GCGCCCTTATACACACAGCAGCTTCTCATGATGATACAGCAGCAGAGAGAGAGCCGATCTC 1349
 OY 661 AAAACCGGCTTACTCTCCAGCCGCGCCGCCCAATCCGACAGCAGCAGCAGCAGCAGCAGCAG 720
 DB 1350 AAAACCGGCTTACTCTCCAGCCGCGCCGCCCAATCCGACAGCAGCAGCAGCAGCAGCAGCAG 1409
 OY 721 TTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 DB 1410 TTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1469
 OY 781 GAGTTTTCGACGCGGAGACTTCTCGAGAGCTGACAGCGGCTGACACAGCAGAGCCTGACG 840
 DB 1470 GAGTTTTCGACGCGGAGACTTCTCGAGAGCTGACAGCGGCTGACACAGCAGAGCCTGACG 1529
 OY 841 AACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 DB 1530 AACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1589
 OY 901 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 DB 1590 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1649
 OY 961 CGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 DB 1650 CGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1709

RESULT 2

AC087298/ 100836 bp DNA linear HTG 20-APR-2002
 LOCUS Homo sapiens chromosome 17 clone RP11-40363 map 17, WORKING DRAFT
 DEFINITION AC087298, 15 ordered pieces.

ACCESSION AC087298.8 GI:20219131
 VERSION AC087298.8 GI:20219131
 KEYWORDS HTG; PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE Homo sapiens.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 100836)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 17, clone RP11-40363
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 100836)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,
 Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Deatellano, K., Dewar, J., Diaz, J.S.,
 Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Harford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Karatas, A., LaRoque, K., Lamzates, R., Landers, T.,
 Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
 McPheters, R., Meldrim, J., Menus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
 Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Submitted (23-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
(bases 1 to 100836)
Birten, B., Linton, L., Nussbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Bastien, Y., Bloom, T., Boguslavsky, L.,
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazarro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S.,
Gard, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Labroque, K., Lamazares, R.,
Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Riedack, M., Riley, R., Risse, C., Rogov, P.,
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 20, 2002 this sequence version replaced g1:20177719.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L12020
Center clone name: 403_G-3
Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 96277 bases at least Q40
Consensus quality: 98051 bases at least Q30
Consensus quality: 98838 bases at least Q20
Insert size: 140000; agarose-fp
Insert size: 99436; sum-of-contigs
Quality coverage: 19.1 in Q20 bases; sum-of-contigs
Quality coverage: 26.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 8096: contig of 8098 bp in length
* 8099 8198: gap of 100 bp
* 8199 8271: contig of 73 bp in length
* 8272 8371: gap of 100 bp
* 8372 9497: contig of 1126 bp in length
* 9498 9597: gap of 100 bp
* 9598 10863: contig of 1266 bp in length
* 10864 10963: gap of 100 bp
* 10964 12232: contig of 1269 bp in length

12233 12332: gap of 100 bp
* 12333 13844: contig of 1512 bp in length
* 13845 13944: gap of 100 bp
* 13945 15215: contig of 1271 bp in length
* 15216 15315: gap of 100 bp
* 15316 18234: contig of 2919 bp in length
* 18235 18334: gap of 100 bp
* 18335 21787: contig of 3453 bp in length
* 21788 21887: gap of 100 bp
* 21888 25485: contig of 3598 bp in length
* 25486 25585: gap of 100 bp
* 25586 40531: contig of 14946 bp in length
* 40532 40631: gap of 100 bp
* 40632 52119: contig of 11488 bp in length
* 52120 52219: gap of 100 bp
* 52220 64941: contig of 12722 bp in length
* 64942 65041: gap of 100 bp
* 65042 79587: contig of 14546 bp in length
* 79588 79687: gap of 100 bp
* 79688 100836: contig of 21149 bp in length.
Location/Qualifiers
1. 100836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-40363"
/clone_11b="RP11-11 Human Male BAC"
1. 8098
/note="assembly-fragment
clone_end:SP6
vector_side:left"
8199..8271
/note="assembly-fragment"
8372..9497
/note="assembly-fragment"
9598..10863
/note="assembly-fragment"
10964..12232
/note="assembly-fragment"
12333..13844
/note="assembly-fragment"
13945..15215
/note="assembly-fragment"
15316..18234
/note="assembly-fragment"
18335..21787
/note="assembly-fragment"
21888..25485
/note="assembly-fragment"
25586..40531
/note="assembly-fragment"
40632..52119
/note="assembly-fragment"
52220..64941
/note="assembly-fragment"
65042..79587
/note="assembly-fragment"
79688..100836
/note="assembly-fragment"

BASE COUNT 23052 a 27082 c 27053 g 2249 t 1400 others
ORIGIN

Query Match 99.78: Score 1076.8; DB 2: Length 100836;
Best Local Similarity 99.88; Pred. No. 2.2e-169;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCAGGCAATCTGTGAGATATCAACACAGCCGCTCAACTAGCACTGACAGT 60
|||||
DB 70785 ATGGCCAGGCAATCTGTGAGATATCAACACAGCCGCTCAACTAGCACTGACAGT 70726
|||||
QY 61 GCTGCTGCTCTCCAGGAAGAGCTGAACCTGAGCGGCCCCAGCGGAGGAGCGGCTG 120
|||||

```

Db 70725 GCTGCTGCTGTCAGGAAGAGCTGAACCTTGAGCCGCCCGGCGGAGAGCGGGTG 70666
QY 121 CCGGAGGAGAGAGTAAAGGCGAATCGAGAGCGTCCCGCAATGGTGGCGCGTCCGGG 180
Db 70665 CCGGAGGAGAGAGTAAAGGCGAATCGAGAGCGTCCCGCAATGGTGGCGCGTCCGGG 70606
QY 181 CCGGAGGAGAGAGTAAAGGCGAATCGAGAGCGTCCCGCAATGGTGGCGCGTCCGGG 240
Db 70605 CCGGAGGAGAGAGTAAAGGCGAATCGAGAGCGTCCCGCAATGGTGGCGCGTCCGGG 70546
QY 241 GAATCTAGCTGCTCTAGAGAGAGGCGGCAAGAGCCGCAATGGGAGCTGTCGCTGGC 300
Db 70545 GAATCTAGCTGCTCTAGAGAGAGGCGGCAAGAGCCGCAATGGGAGCTGTCGCTGGC 70486
QY 301 GCGGACTTCCCGCCCGGCGAGAGTGAACCGGAGCGCGCGAGAGTGTCTGCGCCAG 360
Db 70485 GCGGACTTCCCGCCCGGCGAGAGTGAACCGGAGCGCGCGAGAGTGTCTGCGCCAG 70426
QY 361 CCGTGTCACTGCTCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAG 420
Db 70425 CCGTGTCACTGCTCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAG 70366
QY 421 TGGGAGACAGACAGACAGACAGTGGGGGAGAAAAACATAGAGAGCGCCCTCCAAAGAG 480
Db 70365 TGGGAGACAGACAGACAGACAGTGGGGGAGAAAAACATAGAGAGCGCCCTCCAAAGAG 70306
QY 481 AAGCGCATTTGGAAACCGTACTACAAAGCTGAAGTGGGAGAAAGAAAAAGTTCCAGCAG 540
Db 70305 AAGCGCATTTGGAAACCGTACTACAAAGCTGAAGTGGGAGAAAGAAAAAGTTCCAGCAG 70246
QY 541 AAACAGAGCCTTCGAGCTTCAAGATCCGAGCCGAGATGTTGCCAAGGGCCAGCCGCTC 600
Db 70245 AAACAGAGCCTTCGAGCTTCAAGATCCGAGCCGAGATGTTGCCAAGGGCCAGCCGCTC 70186
QY 601 GCGCCCTATACACACAGCAGTTCCTCATGATGATCAAGACAGAGAGAGCGCGATCTC 660
Db 70185 GCGCCCTATACACACAGCAGTTCCTCATGATGATCAAGACAGAGAGAGCGCGATCTC 70126
QY 661 AAACCGGCGCTTCTACTCCAAAGCGGGCGCGCCCAATCCGAGACACAGCGATGACGAC 720
Db 70125 AAACCGGCGCTTCTACTCCAAAGCGGGCGCGCCCAATCCGAGACACAGCGATGACGAC 70066
QY 721 TTCATAGGAAGAGGGGGTGAAGAGATGGGGGACAGATGGGATGGGAGGGAGCGGCAC 780
Db 70065 TTCATAGGAAGAGGGGGTGAAGAGATGGGGGACAGATGGGATGGGAGGGAGCGGCAC 70006
QY 781 GAGTTTCTGACGCGGAGCTTCTCGGAGAGCTACGAGCGGTACACACAGAGAGCTGACG 840
Db 70005 GAGTTTCTGACGCGGAGCTTCTCGGAGAGCTACGAGCGGTACACACAGAGAGCTGACG 69946
QY 841 AACATGAGCAGCAGAGCTCATCAAGAGTACTGGAAGCTGGAGAACTGCTCTGCGCG 900
Db 69945 AACATGAGCAGCAGAGCTCATCAAGAGTACTGGAAGCTGGAGAACTGCTCTGCGCG 69886
QY 901 AAGGAGAGAGAGAGAGAGCGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Db 69885 AAGGAGAGAGAGAGAGAGCGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69826
QY 961 GGTGTGCGGAGCTGAGAGTGAAGTGAACCGGCTGCGCGCGAGAACTCCAGCTGTG 1020
Db 69825 GGTGTGCGGAGCTGAGAGTGAAGTGAACCGGCTGCGCGCGAGAACTCCAGCTGTG 69766
QY 1021 ACCGAGAGAGAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 69765 ACCGAGAGAGAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69706

```

RESULT 3
 AK023624 2178 bp mRNA linear pri 01-AUG-2002
 LOCUS Home sapiens cDNA FLJ13562 fis, clone PLACE1008080, highly similar
 DEFINITION to Homo sapiens mRNA for HEXIM1 protein.
 ACCESSION AK023624

```

VERSION AK023624.1 GI:10435606
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens Placenta cDNA to mRNA, clone_1lib.PLACE1
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
    Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
    Magatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
    Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
    Ono,Y., Takiguchi,S., Matanabe,S., Kimura,K., Murakami,K.,
    Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
    Nakamura,Y., Nagahara,K., Masuhio,Y., Ninomiya,K. and Iwayanagi,T.
    NEDO human cDNA sequencing project
  2 (bases 1 to 2178)
  Unpublished
REFERENCE
  1 Isogai,T. and Otsuki,T.
    Direct Submission
    Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
    Genomics Laboratory: 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
    (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT
  NEDO human cDNA sequencing project supported by Ministry of
  International Trade and Industry of Japan; cDNA full insert
  sequencing: Research Association for Biotechnology; cDNA library
  construction, 5'- & 3'-end one pass sequencing and clone selection:
  Helix Research Institute (supported by Japan Key Technology Center
  etc.) and Department of Virology, Institute of Medical Science,
  University of Tokyo.
FEATURES
  source
    1..2178
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="PLACE1008080"
    /tissue.type="Placenta"
    /clone_1lib="PLACE1"
    /note="Cloning vector: pME18SFL3"
BASE COUNT
  525 a 513 c 678 g 462 t
ORIGIN
  Query Match 99.6%; Score 1075.2; DB 9; Length 2178;
  Best Local Similarity 99.7%; Pred. No. 6.5e-169;
  Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
  1 ATGCCGAGGCCATTCTGTTCAGAAATATCAACACGACCTCAAACTAGCACTGTACAGT 60
  716 ATGGCCGAGGCCATTCTGTTCAGAAATATCAACACGACCTCAAACTAGCACTGTACAGT 775
  61 GCTGCTGCTGTCCAGAGAGCTGAACCTTGAGCCGCCCGGAGGGCGGAGAGAGGGGTG 120
  776 GCTGCTGCTGTCCAGAGAGCTGAACCTTGAGCCGCCCGGAGGGCGGAGAGAGGGGTG 835
  121 CCCGAGAGAGCAGTAGGTGCAATCGAGAGAGCTTCCCGCAATGGTGGTCCGCGGG 180
  836 CCCGAGAGAGCAGTAGTAGTGGCAATCGAGAGAGCTTCCCGCAATGGTGGTGGTCCG 895
  181 CCGAGAGGGGAGAGAGCGCTGGAATCCCAACCACTCTCTTCAGACACCGAGCGCTGTCCA 240
  896 CCGAGAGGGGAGAGAGCGCTGGAATCCCAACCACTCTCTTCAGACACCGAGCGCTGTCCA 955
  241 GAATCTAGCTGCTCTAGAGAGAGGGGCGAAGAGGGGCGAAGTGGGAGAGAGCTGCTGG 300
  956 GAATCTAGCTGCTCTAGAGAGAGGGGCGAAGAGGGGCGAAGTGGGAGAGAGCTGCTGG 1015
  1016 GCGGACTTCCCGCCCGGCGAGAACTGGAACCGAGAGCGCGAGAGAGAGAGAGAGAG 1075
  361 CCGTGTCACTGCTCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAGAG 420
  1076 CCGTGTCACTGCTCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAGAG 1135

```

```

OY 421 TGGGACAGACAGACAGACAGCTGGGGAGAAAAACATTAAGAGACGCCCTCCAGAAG 480
    |||||||
Db 1136 TGGGACAGACAGACAGACAGCTGGGGAGAAAAACATTAAGAGACGCCCTCCAGAAG 1195
OY 481 AAGCGCATTTGGAAACCTTACTACAAAGCTGAACCTGGGAAGAGAAAAAGTTCCAGCAG 540
    |||||||
Db 1196 AAGCGCATTTGGAAACCTTACTACAAAGCTGAACCTGGGAAGAGAAAAAGTTCCAGCAG 1255
OY 541 AACAGACCTTTCAGAGCTTCAAGGATCCGACCCGAGATGTTGCCCAAGGCCAGCCGTC 600
    |||||||
Db 1256 AACAGACCTTTCAGAGCTTCAAGGATCCGACCCGAGATGTTGCCCAAGGCCAGCCGTC 1315
OY 601 GCGCCCTTAAACACACAGCAGTTCCTCATGTGATGATCAGACACAGAGAGAGCCGATCTC 660
    |||||||
Db 1316 GCGCCCTTAAACACACAGCAGTTCCTCATGTGATGATCAGACACAGAGAGAGCCGATCTC 1375
OY 661 AAAACGGCTGTACTCCAGCGGGCCGCCCAATCCGACGACACAGCAGGATGACAGC 720
    |||||||
Db 1376 AAAACGGCTGTACTCCAGCGGGCCGCCCAATCCGACGACACAGCAGGATGACAGC 1435
OY 721 TTTCATGGAAGAAGGGGGTGAGAGAGATGGGGGACCCGATGGATGGAGAGCGCAGC 780
    |||||||
Db 1436 TTTCATGGAAGAAGGGGGTGAGAGAGATGGGGGACCCGATGGATGGAGAGCGCAGC 1495
OY 781 GAGTTTTCGACGGGAGCTTCTCGAGAGAGTACGAGCGGTACACACAGAGAGAGCCTGCGAG 840
    |||||||
Db 1496 GAGTTTTCGACGGGAGCTTCTCGAGAGAGTACGAGCGGTACACACAGAGAGAGCCTGCGAG 1555
OY 841 AACATGACCAAGCAGAGAGCTCATCAAGAGAGTACCTGGAATCGAGAACTGCTCTCGCGC 900
    |||||||
Db 1556 AACATGACCAAGCAGAGAGCTCATCAAGAGAGTACCTGGAATCGAGAACTGCTCTCGCGC 1615
OY 901 ATGGAGACGGAACAAACGGCTGGCGCTGGAGAGCAACAGCGCTGGTGGCAGACAGCG 960
    |||||||
Db 1616 ATGGAGACGGAACAAACGGCTGGCGCTGGAGAGCAACAGCGCTGGTGGCAGACAGCG 1675
OY 961 CGTGTCCGAGAGCTGGAGAGCTGGAGACCGGCTGGCGCGCGGAGAACCTCCAGCTGCTG 1020
    |||||||
Db 1676 CGTGTCCGAGAGCTGGAGAGCTGGAGACCGGCTGGCGCGCGGAGAACCTCCAGCTGCTG 1735
OY 1021 ACCGAGAACGAACCTGACACCGGAGAGAGAGAGCGCGCTTCCCAAGTTTGGAGACTAG 1080
    |||||||
Db 1736 ACCGAGAACGAACCTGACACCGGAGAGAGAGAGAGCGCGCTTCCCAAGTTTGGAGACTAG 1795

RESULT 4
BC006460 2086 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, HMBA-Inducible, clone MGC:1880 IMAGE:353529, mRNA,
DEFINITION complete cds.
ACCESSION BC006460
VERSION BC006460.1 GI:13623668
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2086)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
BC Cancer Agency, Vancouver, BC, Canada
Info@bcgsc.bc.ca

```

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Meta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carlie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Ilija Prabh, Parvaneh Seedei, Jacqueline Schein, Duane Smalish, Michael Tsai, Loraine Speede, Jeff Stolt, Michael Thorne, Miranada Tsai, Natalja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL, Plate: 8 Row: d Column: 1
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4062855.

FEATURES

source

1..2086

/organism="Homo sapiens"
 /db_xref="locusid:10614"
 /db_xref="taxon:9606"
 /clone="MGC:1880 IMAGE:353529"
 /tissue_type="Lung, small cell carcinoma"
 /clone_id="NIH-MGC_7"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 606..1685
 /codon_start=1
 /product="HMBA-Inducible"
 /protein_id="AAH06460.1"
 /db_xref="gi:13623669"

CDS

/translation="MAEPFLSEYHOPQTSNCTGAAYVOEELNPERPGEAREVEED
 SRMOSRAFPOLGRRPGEPEESLQSPPLTOGCPRESSCJREBEKGNDSAGGD
 FPPAEVETPEPEELIAPQCHDSFSAKAGAAAGGEEGGOOLDKKKRRRPSK
 KRHKPKYKLWEKKKTKDEQSLRASHIRAMPAKGPVAPYNTQITLMDHOEE
 DKTGLSKRAAKSDTSDDFMEGGEEDGSGKGGDSERLQRFSTTERHT
 ESLDNMSKDELIKYLELEKLSRHEEDNNRLRLESKRLGGDARVRELELDRLRA
 ENLQLTENEHLRQERAPLSKFD"
 BASE COUNT 510 a 505 c 618 g 453 t
 ORIGIN

Query Match

Best Local Similarity 99.4%; Score 1073.6; DB 9; Length 2086;

Matches 1076; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

OY 1 ATGGCGGAGCCATTGTCAGAAATATACACAGCCTCAACTAGCACTGTACAGT 60
    |||||||
Db 606 ATGGCGGAGCCATTGTCAGAAATATATACACAGCCTCAACTAGCACTGTACAGT 665
OY 61 GCTGCTGTCGCCAGGAAGAGCTGAAACCTTAGAGCGCCGCCAGGCGGAGAGCGGTG 120
    |||||||
Db 666 GCTGCTGTCGCCAGGAAGAGCTGAAACCTTAGAGCGCCGCCAGGCGGAGAGCGGTG 725
OY 121 CCCGAGAGAGACAGTAGTGGCAATCGAGACGCTTCCGCCAGTTGGGTGGCCGCGGG 180
    |||||||
Db 666 GCTGCTGTCGCCAGGAAGAGCTGAAACCTTAGAGCGCCGCCAGGCGGAGAGCGGTG 725
OY 121 CCCGAGAGAGACAGTAGTGGCAATCGAGACGCTTCCGCCAGTTGGGTGGCCGCGGG 180
    |||||||
Db 726 CCCGAGAGAGACAGTAGTGGCAATCGAGACGCTTCCGCCAGTTGGGTGGCCGCGGG 785
OY 726 CCCGAGAGAGACAGTAGTGGCAATCGAGACGCTTCCGCCAGTTGGGTGGCCGCGGG 785
    |||||||
Db 181 CCGGAGGGGGAAGGAGGCTGGAATCCCAACCACTGCTTGGAGACCCAGGCGGTGCCA 240
    |||||||
Db 786 CCGGAGGGGGAAGGAGGCTGGAATCCCAACCACTGCTTGGAGACCCAGGCGGTGCCA 845
OY 241 GAATCTACCTGCTTCAGAGAGAGCGGAGAGGCGCGCAATGGGAGAGACTCTGCCCTGGC 300
    |||||||
Db 846 GAATCTACCTGCTTCAGAGAGAGCGGAGAGGCGCGCAATGGGAGAGACTCTGCCCTGGC 905
OY 301 GCGCAGCTTCCGCGCGCGGCGGAGAGTGAACCGCGCGCGGAGCGGAGCTGCTGCCAG 360
    |||||||
Db 906 GCGCAGCTTCCGCGCGCGGCGGAGAGTGAACCGCGCGCGGAGCGGAGCTGCTGCCAG 965
OY 361 CCTGTCTATGACTCCGAGAGCGCAGTAAGTTGGGGGCTCTGGCGCAGGGGAGAGAGAG 420
    |||||||
Db 966 CCTGTCTATGACTCCGAGAGCGCAGTAAGTTGGGGGCTCTGGCGCAGGGGAGAGAGAG 1025
OY 421 TGGGACAGCAGACAGACAGCTGGGGAGAAAAACATTAAGAGACGCCCTCCAGAAG 480
    |||||||

```

```

Db      1026 TGGGAGACGACGACGAGCTGGGGAGAGAAAAAATAGAGACGCCGTCAGAGAG 1085
QY      481 AAGCGCATTTGAAACCGTACTACTAAGCTGAACTGGGAGAGAGAAAAAGTTGACGAG 540
Db      1086 AAGCGCATTTGAAACCGTACTACTAAGCTGAACTGGGAGAGAGAAAAAGTTGACGAG 1145
QY      541 AAGCAGACCTTTCAGCTTCAGAGATCCGAGCCGAGATGTTGCCAAGGGCCAGCCGCTC 600
Db      1146 AAGCAGACCTTTCAGCTTCAGAGATCCGAGCCGAGATGTTGCCAAGGGCCAGCCGCTC 1205
QY      601 GCGCCCTATTAACACCGAGTTCCTCATGATGATCAGACGAGGAGGAGCCGAGCTC 660
Db      1206 GCGCCCTATTAACACCGAGTTCCTCATGATGATCAGACGAGGAGGAGCCGAGCTC 1265
QY      661 AAGCAGACCTTTCAGCTTCAGAGATCCGAGCCGAGATGTTGCCAAGGGCCAGCCGCTC 720
Db      1266 AAGCAGACCTTTCAGCTTCAGAGATCCGAGCCGAGATGTTGCCAAGGGCCAGCCGCTC 1325
QY      721 TTCATGGAAGAGGGGGTGAAGAGATGGGGGACGAGTGGATGGAGGGGACGCGAGC 780
Db      1326 TTCATGGAAGAGGGGGTGAAGAGATGGGGGACGAGTGGATGGAGGGGACGCGAGC 1385
QY      781 GAGTTTCTGACGAGGAGCTTCGAGACGTACGAGGAGGAGTACCAACAGGAGACCTGACG 840
Db      1386 GAGTTTCTGACGAGGAGCTTCGAGACGTACGAGGAGGAGTACCAACAGGAGACCTGACG 1445
QY      841 AACATGAGCAGAGGAGAGCTCATCAAGAGTACCTGGAATGGAAGAGTCCCTCTGCGCG 900
Db      1446 AACATGAGCAGAGGAGAGCTCATCAAGAGTACCTGGAATGGAAGAGTCCCTCTGCGCG 1505
QY      901 ATGAGAGAGAGAGAACCAACCGGCTGCGGCTTGAGAGACAGGCGCTGGGTGGCGACGAGCG 960
Db      1506 ATGAGAGAGAGAGAACCAACCGGCTGCGGCTTGAGAGACAGGCGCTGGGTGGCGACGAGCG 1565
QY      961 CGTGTGCGGAGAGTGAAGCTGAGCTGAGACCGGCTGCGGCGCCGAGAACTCCAGCTGCTG 1020
Db      1566 CGTGTGCGGAGAGTGAAGCTGAGCTGAGACCGGCTGCGGCGCCGAGAACTCCAGCTGCTG 1025
QY      1021 ACCGAGACGAGTGCACCGGCGAGCAGAGCAGCGCCGCTTTCCAAGTTTGAGAGCTAG 1080
Db      1626 ACCGAGACGAGTGCACCGGCGAGCAGAGCAGCGCCGCTTTCCAAGTTTGAGAGCTAG 1685

RESULT 5
AC024047/c 142326 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 17 clone RP11-403G3, WORKING DRAFT
DEFINITION
AC024047
AC024047.3 GI:8570008
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 142326)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 142326)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2000) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Jun 16, 2000 this sequence version replaced gi:7109599.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0403603

```

```

----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: Plasmid, 0%
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 127942 bases at least Q40
Consensus quality: 13308 bases at least Q30
Consensus quality: 135734 bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 139626; sum-of-coverage
Quality coverage: 4.06 in Q20 bases; agarose-fp
Quality coverage: 4.33 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1575: contig of 1575 bp in length
1576 1675: gap of unknown length
1676 3114: contig of 1438 bp in length
3115 3214: gap of unknown length
3215 5389: contig of 2175 bp in length
5390 5489: gap of unknown length
5490 7129: contig of 1640 bp in length
7130 7229: gap of unknown length
7230 9240: contig of 2011 bp in length
9241 9340: gap of unknown length
9341 11177: contig of 1837 bp in length
11178 11277: gap of unknown length
11278 13113: contig of 1836 bp in length
13114 13213: gap of unknown length
13214 16526: contig of 3313 bp in length
16527 16626: gap of unknown length
16627 19864: contig of 3238 bp in length
19865 19964: gap of unknown length
19965 22667: contig of 2703 bp in length
22668 22767: gap of unknown length
22768 26550: contig of 3883 bp in length
26551 26750: gap of unknown length
26751 29303: contig of 2553 bp in length
29304 32073: contig of 2670 bp in length
32074 32173: gap of unknown length
32174 37141: contig of 4968 bp in length
37142 37241: gap of unknown length
37242 40838: contig of 3497 bp in length
40839 45166: gap of unknown length
45167 45266: contig of 4328 bp in length
45267 49501: gap of unknown length
49502 49601: contig of 4235 bp in length
49602 55647: gap of unknown length
55648 55747: contig of 6046 bp in length
55748 61532: contig of 5785 bp in length
61533 61632: gap of unknown length
61633 68105: contig of 6373 bp in length
68106 68105: gap of unknown length
68106 75211: contig of 7106 bp in length
75212 81560: contig of 6249 bp in length
81561 81660: gap of unknown length
81661 88524: contig of 6864 bp in length
88525 88624: gap of unknown length
88625 95519: contig of 6895 bp in length
95520 105148: gap of unknown length
105149 105248: contig of 9529 bp in length
105249 115269: contig of 10021 bp in length

```


FEATURES
source
1. 142326
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-403g3"
BASE COUNT 32503 a 37098 c 37256 g 32736 t 2733 others
ORIGIN

Query Match 97.5%: Score 1052.8; DB 2; Length 142326;
Best Local Similarity 99.6%: Pred. No. 2e-165;
Matches 1076; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 ATGGCCGAGCCATCTTGTCAAGATATCAACACCAGCTCAAACTAGCACTGACAGT 60
|||
Db 108505 ATGGCCGAGCCATCTTGTCAAGATATCAACACCAGCTCAAACTAGCACTGACAGT 108446
|||

QY 61 GCTCTGCTGTCCAGGAAGAGCTAACCCTGAGCGCCCGCAGCGCGAGAGCGGCTG 120
|||
Db 108445 GCTCTGCTGTCCAGGAAGAGCTAACCCTGAGCGCCCGCAGCGCGAGAGCGGCTG 108386
|||

QY 121 CCCGAGGAGCACTAGTGGCAATCGAGAGCTTCCCGCCAGTGGGGGCGCTCCGCGG 180
|||
Db 108385 CCCGAGGAGCACTAGTGGCAATCGAGAGCTTCCCGCCAGTGGGGGCGCTCCGCGG 108326
|||

QY 181 CCGGAGGAGGGAAGGAGCTGGAATCCCAACACCTCCCTTGCAGAGCCAGCGCTGTCCA 240
|||
Db 108325 CCGGAGGAGGGAAGGAGCTGGAATCCCAACACCTCCCTTGCAGAGCCAGCGCTGTCCA 108266
|||

QY 241 GAATCTAGCTCCCTGAGAGAGGCGGAGAGAGGCCAAGATGGGGACGACTCTCCGCTGCGC 300
|||
Db 108265 GAATCTAGCTCCCTGAGAGAGGCGGAGAGAGGCCAAGATGGGGACGACTCTCCGCTGCGC 108206
|||

QY 301 GGCAGCTTCCCGCGCGCGAGAGAGTGAACCGAGCGCGCGCGAGCTCTGCGCCAG 360
|||
Db 108205 -GCGACTTCCCG -CGCGCGCGAGAGAGTGAACCGAGCGCGCGAGCTCTGCGCCAG 108148
|||

QY 361 CCTTGTGATGACTCCGAGGCGAGTGAAGTTGGGGGCTCTCCCGCGAGGGGGGGAAGAGAG 420
|||
Db 108147 CCTTGTGATGACTCCGAGGCGAGTGAAGTTGGGGGCTCTCCCGCGAGGGGGGGAAGAGAG 108088
|||

QY 421 TGGGGAAGAGCGAGAGAGAGCTGGGGAAGAAAAAACAATAAGAGACCGCGCTCCAAGAG 480
|||
Db 108087 TGGGGAAGAGCGAGAGAGAGCTGGGGAAGAAAAAACAATAAGAGACCGCGCTCCAAGAG 108028
|||

QY 481 AAGGGGATTGGAACCGTACTACAGTGAAGTGGGGAAGAAAAAGTTGAGAGAG 540
|||
Db 108027 AAGGGGATTGGAACCGTACTACAGTGAAGTGGGGAAGAAAAAGTTGAGAGAG 107968
|||

QY 541 AAACAGAGCCTTCGAGCTTCAAGGATCCAGCCGAGATGTTCCGCAAGGGCGCACCGGTC 600
|||
Db 107967 AAACAGAGCCTTCGAGCTTCAAGGATCCAGCCGAGATGTTCCGCAAGGGCGCACCGGTC 107908
|||

QY 601 GCGGCCATTAACACACAGAGATTCTCATGATATATACAGACCGAGAGAGCCGGATCTC 660
|||
Db 107907 GCGGCCATTAACACACAGAGATTCTCATGATATATACAGACCGAGAGAGCCGGATCTC 107848
|||

QY 661 AAAACCGGCTGTACTCTCAAGCGGGCGCGCCAAATCCGAGCAGACAGAGGATGAGAGAC 720
|||
Db 107847 AAAACCGGCTGTACTCTCAAGCGGGCGCGCCAAATCCGAGCAGACAGAGGATGAGAGAC 107788
|||

QY 721 TTCTATGGAAGAGGGGTGAGAGAGATGGGGGAGCGCATGGAGTGGAGGGGAGCGGAGC 780
|||
Db 107787 TTCTATGGAAGAGGGGTGAGAGAGATGGGGGAGCGCATGGAGTGGAGGGGAGCGGAGC 107728
|||

QY 781 GAGTTTGTGAGCGCGGACTTCTGCGAGACGTAGAGCGGATACCAACAGGAGAGAGCCTGACG 840
|||
Db 107727 GAGTTTGTGAGCGCGGACTTCTGCGAGACGTAGAGCGGATACCAACAGGAGAGAGCCTGACG 107668
|||

QY 841 AACATGAGCAGCAGGAGCTCATCAGAGTACCTGGAGTGGAGAGTGCCTCCGCC 900
|||
Db 107667 AACATGAGCAGCAGGAGCTCATCAGAGTACCTGGAGTGGAGAGTGCCTCCGCC 107608
|||

QY 901 ATGAGAGCAGAGAACCAACCGGCTCGGCTGGAGAGCAAGCGGCTGGGAGAGCAGCGC 960
|||
Db 107607 ATGAGAGCAGAGAACCAACCGGCTCGGCTGGAGAGCAAGCGGCTGGGAGAGCAGCGC 107548
|||

QY 961 CGTGTGCGGAGAGCTGAGAGCTGAGACCGGCTCGCGCGCGAGAACCTCCAGCTGCTG 1020
|||
Db 107547 CGTGTGCGGAGAGCTGAGAGCTGAGACCGGCTCGCGCGCGAGAACCTCCAGCTGCTG 107488
|||

QY 1021 ACCGAGAACGAACTGCACCGCGACAGAGAGCGCGCTTCCTCAACTTGGAGACTAG 1080
|||
Db 107487 ACCGAGAACGAACTGCACCGCGACAGAGAGCGCGCTTCCTCAAGTTGGAGACTAG 107428
|||

RESULT 6
LOCUS AY090614 3402 bp mRNA linear ROD 18-JUL-2002
DEFINITION Mus musculus cardiac lineage protein 1 (Clp1) mRNA, complete cds.
ACCESSION AY090614
VERSION AY090614.1 GI:20135640
KEYWORDS
SOURCE
ORGANISM Mus musculus.
MUS musculus.
Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3402)
Huang, F., Wagner, M. and Siddiqui, M. A. O.
Structure, expression, and functional characterization of the mouse
Clp-1 gene
JOURNAL Gene 292 (1-2), 245-259 (2002)
MEDLINE 22114986
PUBMED 12119119
REFERENCE 2 (bases 1 to 3402)
Huang, F., Wagner, M. and Siddiqui, M.
Direct Submission
AUTHORS Submitted (19-MAR-2002) Anatomy and Cell Biology, State University
of New York Health Science Center, 450 Clarkson Avenue, Brooklyn,
NY 11203, USA
FEATURES
source
Location/Qualifiers
1. 3402
/organism="Mus musculus"
/strain="129/SvJ"
/db_xref="taxon:10090"
1. 3402
/gene="Clp1"
596..1666
/gene="Clp1"
/note="nuclear protein; CLP-1"
/codon_start=1
/product="cardiac lineage protein 1"
/protein_id="AA09026.1"
/translation="MAEPLLTLEHODHOTSNCNGAAYVHEHTSERPPSAEPPPKED
SRMSRSASLDGSRPOEGEGKGLKHLPIPTLNACPLSLBKEGKONKEDISTGSA
SPSAEGEPMSKESLVPEGHDEATKQEAIPAAGGEPPWQOQRLKKHRRPSSKKRH
WKPYKLTWEKKEKFKDEKSLIRASRYAEKFAKQVPAPYNTQFLMDHDOEPDLK
TGLYPKRAAKSDTSDDEVEEAGEEDGSDGSGGSEFLQRFSEYERYHAESTL
ONMSKLEIKEYLELEKLSRKEDENNRLRLSKRLGGVDARVRELELEIDRURENL
QLLTENLHQOERAPLTKFGD"
BASE COUNT 920 a 771 c 890 g 821 t
ORIGIN

Query Match 77.9%: Score 840.8; DB 10; Length 3402;
Best Local Similarity 86.9%: Pred. No. 4.7e-130;
Matches 939; Conservative 0; Mismatches 132; Indels 9; Gaps 1;

QY 1 ATGGCCGAGCCATCTTGTCAAGATATCAACACCAGCTCAAACTAGCACTGACAGT 60
|||
Db 596 ATGGCCGAGCCATCTTGTCAAGATATCAACACCAGCTCAAACTAGCACTGACAGT 655
|||

```

OY 61 GCTGCTGTCACAGAGAGCTGAACCTTGACGCCGCCAGGCGGAGAGCGGGTG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 656 GCTGCTGTCACAGAGAGCTGAACCTTGACGCCGCCAGGCGGAGAGCGGGTG 715
OY 121 CCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 716 CCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 775
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 181 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 776 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 835
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 241 GAATCTAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 836 GAATGAGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 301 GCGGACCTTCCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 896 GGTGCTCTCCGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 361 CTTTCTATGACTCCGAGGCGAGTAAAGTTGGGGGCTCTGCGCGAGGCGAGAGAG 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 947 CCAGGTCAATGACTCCGAGGCGAGTAAAGTTGGGGGCTCTGCGCGAGGCGAGAGAG 1006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 421 TGGGAGACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1007 TGGGAGACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1066
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 481 AAGCGGCAATGGAACCGTACTACAAAGTGAACCTGGGAAGAGAGAGAGAGAG 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1067 AAGCGGCAATGGAACCGTACTACAAAGTGAACCTGGGAAGAGAGAGAGAGAG 1126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 541 AAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1127 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 601 GCGGCTCTATACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1187 GCGGCTCTATACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 661 AAACCGGCTCTACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1247 AAACCGGCTCTACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 721 TTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1307 TTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 781 GAGTTTTCGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1367 GAGTTTTCGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 841 AACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1427 AACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 901 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1487 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 961 CGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1547 CGAGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1021 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1607 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 7
AL731805 111002 bp DNA linear ROD 21-JUN-2002
LOCUS AL731805
DEFINITION Mouse DNA sequence from clone RP23-358E19 on chromosome 11,

complete sequence.
AL731805
VERSION AL731805.8 GI:21540125
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 111002)
AUTHORS Dunn, M.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Jun 23, 2002 this sequence version replaced gi:21531483.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M3 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> RP23-358E19 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6

FEATURES

source

1..111002
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-358E19"
/clone_11b="RPI-23"
BASE COUNT 28051 a 27579 c 27135 g 28237 t
ORIGIN

Query Match 77.9%; Score 840.8; DB 10; Length 111002;
Best Local Similarity 86.9%; Pred. No. 3e-130;
Matches 939; Conservative 0; Mismatches 132; Indels 9; Gaps 1;

```

OY 1 ATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91440 ATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91499
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 61 GCTGCTGTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91500 GCTGCTGTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91559
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 121 CCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91560 CCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91619
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 181 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91620 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91679
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

OY	241	GAATCTAGCTGCTCTGAGAGAGGGCGAGAAAGGGCCAGAAATGGGGAGCACTGTCCGTGGC	300
Db	91680	GAATTTGACTTCCTCCCTGGAGAAAGGGCGAGAAAGGGCCAGAAATGGGGAGCACTTATCCACTGTGC	91739
OY	301	GGCAGCACTTCCCGCCGGCGGCGAGAAAGTGAACCCGAGCCCGAGGCGGAGCTCGGCCGAG	360
Db	91740	GGTCCCTCCCCGTCGTGGCGGAGGAGGA-----GCCCATGTTCAGAGTCCCTGTGCGAG	91790
OY	361	CCCTTGTGATGACTCCGAGGGCCAGTAAATTGGGGGCTCTGTCCGACGAGGGGCGAGAGGAG	420
Db	91791	CCAGGTGATGACTCGAGGGCGACACCAAGACGAGAGGCTCTGTCCGCGGAGGCGAGAGCCA	91850
OY	421	TGGGGACACGACGACGACAGCTGTGGGAGAGAAAAAATPAGAGACGCCCTTCGAAGAG	480
Db	91851	TGGGGACACGACGACGACAGCTGTGGGAAACAAAAAATCTGAGAGAGCCCTCTCAAAAGAG	91910
OY	481	AAGCGGCAATTGGAAACCGCTACTCTCAAGCTGAACTGGGAGAGAGAAAAAAGTTCTGACGAG	540
Db	91911	AAGCGGCAATTGGAAAGCCCTACTACTACAACTGACTTTGGGAGAGAGAAAAAAGTTCTGACGAG	91970
OY	541	AAACGAGACCTTGGAGCTTCAAGGATTCGAGCCGAGATGTTGCCCAAGGGCCAGCCCGTTC	600
Db	91971	AAGCAGAGCCTGCGAGCTTGGCGGGTTCGAGCCGAGATGTTGCCCAAGGGCCAGCCAGTT	92030
OY	601	GGCGCCCTATAACACACGACAGTTCCTATGTGATGATCAGACACGAGAGAGCCGGATCTTC	660
Db	92031	GGCGCCCTATAACACACGACAGTTCCTATGTGATGATGATCAGACACGAGAGAGCCGTGATCTTC	92090
OY	661	AAACCCGGCCTGTACTCTCAAGCGGGCGCCGCCCAATTCGACGACACCCAGCCGATGACGAC	720
Db	92091	AAACCCGGCCTTTTACCCCAAGCGGGCGACCCGCCCAATTCGACGACACCCAGCCGATGAGGAT	92150
OY	721	TTTCATGGAGAGAGGGGGTGTAGAGAGATGGGGGCGACGCGATGGATGGGAGGGGACGGCACG	780
Db	92151	TTTGTGGAGAGAGCTGCTGAGAGAAAGACGGAGCGACGAGATGGATGGGAGGGGACGGCACG	92210
OY	781	GAGTTTCTGCAAGCGGGACTTCTGGAGAGCTACGAGCGGTACACACGGAGAGCCTGTGAG	840
Db	92211	GAGTTTCTGCAAGCGGGACTTCTGGAGAGCCTACGAGGGGTACACCCCGAGAGCCTGTGAG	92270
OY	841	AACATGAGCAAGCAGGAGCTCATCAACAGGACTTACCTGGAACTGGAGAAAGTCCCTGTGGCG	900
Db	92271	AACATGAGCAAGCAGGAGCTCATCAACAGGACTTACCTGGAGAGTGGAGAAAGTCCCTGTCCGCG	92330
OY	901	ATGAGAGGACGAGAAACAACCGGCTGCGGCTGGAGAGCAACCGGCTGGGTGGCGAGACGGG	960
Db	92331	AAGGAGGACGAAATATACGGGGCTGGCGCTGGAAAGCAAGCGGGCTGGGAGGGCGTTCGACGGG	92390
OY	961	CGTCTGGGGAGCTGTGAGCTGTGAGCTGTGACCGGCTGGCGCGGCGAGAAACCTCCAGCTGTCTG	1020
Db	92391	CGACTGGGGAGCTGTGAGCTGTGAGCTGTGAGCTGTGACCGGCTGGCGCGCTCCAGCACTGTCTG	92450
OY	1021	ACCGAGAAAGCAACTGCAACCGGCGACGAGACGAGCGCGCTTTCCAAGTTTGGAGACTAG	1080
Db	92451	ACCGAGAAAGCAACTGCAACCGGCGACGAGACGAGCGCGCTTTCCAAGTTTGGAGACTAG	92510
RESULT 8			
LOCUS	BC022111	1488 bp	linear
DEFINITION	Mus musculus, similar to HMB-Inducible, clone MGC:36249		
ACCESSION	BC022111		
KEYWORDS	BC022111.1	GI:18381039	
SOURCE	MGC.		
ORGANISM	house mouse.		
	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	Strausberg, R.		
AUTHORS	Direct Submission		
TITLE	Submitted (24-JAN-2002)		
JOURNAL	National Institutes of Health, Mammalian		

REMARK	COMMENT
NIH-MGC Project URL: http://mgc.ncl.nih.gov	Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
Contact: MGC help desk	NIH-MGC Project URL: http://mgc.ncl.nih.gov
Email: cgapbs-remail.nih.gov	NIH-MGC Project URL: http://mgc.ncl.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.	NIH-MGC Project URL: http://mgc.ncl.nih.gov
cDNA Library Preparation: Life Technologies, Inc.	NIH-MGC Project URL: http://mgc.ncl.nih.gov
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)	NIH-MGC Project URL: http://mgc.ncl.nih.gov
DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada	NIH-MGC Project URL: http://mgc.ncl.nih.gov
info@bgsc.bc.ca	NIH-MGC Project URL: http://mgc.ncl.nih.gov
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chin, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Seedei, Jacqueline Schein, Diane Smalls, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.	NIH-MGC Project URL: http://mgc.ncl.nih.gov
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov	NIH-MGC Project URL: http://mgc.ncl.nih.gov
Series: IRAX Plate: 57 Row: A Column: 15	NIH-MGC Project URL: http://mgc.ncl.nih.gov
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.	NIH-MGC Project URL: http://mgc.ncl.nih.gov
Location/Qualifiers	NIH-MGC Project URL: http://mgc.ncl.nih.gov
1. 1488	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/organism="Mus musculus"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/db_xref="taxon:10090"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/map="C2ECH 11"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/clone="MGC:36249 IMAGE:5067620"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/clone_lib="NCL-CGAP-Lu29"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/lab_host="DH10B"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/note="Vector: pcMV-SPORT6"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
448. 1188	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/codon_start=1	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/product="Similar to HmBA-Inducible"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/protein_id="AAH2111.1"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/db_xref="GI:18381040"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
/translation="MSESLVPGHDSSEATKQEAFAAGGSEPPWGOOROLGKKHRRRR SKRRHMKPYKLLMEKKKFKDESKLASVRAEMFAKQCPVAPYNTOTPLMDHDO EEPRLKTLVPRRAAKSDPDSDFVFAEGEEDGSGMGDSSEFLQRFSTYER EYHASELWMSKQELIKETLELEKLSKREDDNNRLRLSKRLGVDAVRLELELDR LRAEHLDTLENEHLRQDERAPLSKFGD"	NIH-MGC Project URL: http://mgc.ncl.nih.gov
BASE COUNT	NIH-MGC Project URL: http://mgc.ncl.nih.gov
ORIGIN	NIH-MGC Project URL: http://mgc.ncl.nih.gov
Query Match	NIH-MGC Project URL: http://mgc.ncl.nih.gov
Best Local Similarity 86.9%; Pred. No. 7.5e-129;	NIH-MGC Project URL: http://mgc.ncl.nih.gov
Matches 932; Conservative 0; Mismatches 132; Indels 9; Gaps 1;	NIH-MGC Project URL: http://mgc.ncl.nih.gov
8 AGCCATTCTTTCAGCAATATATCAACACAGCCTCAATAGCAACTGTACAGTCTGCTG 67	NIH-MGC Project URL: http://mgc.ncl.nih.gov
125 AGCCACTTTCAGCAGACATCAACACAGCCTCAATAGCAACTGTACAGTCTGCTG 184	NIH-MGC Project URL: http://mgc.ncl.nih.gov
68 CTGTCAGGAAGAGCTGAACCTGAGCGCCGCCAGCGCGAGAGAGCGGTGCCGAGG 127	NIH-MGC Project URL: http://mgc.ncl.nih.gov
185 TTGTTCATGAAAGAGATACCTCTGTAGCGCCGCCCAAGCGCGAGAAAGGTTGCCAAGG 244	NIH-MGC Project URL: http://mgc.ncl.nih.gov
128 AGGACAGTAGTGGCAATTCGAGAGCGTTTCCCCCACTTGGTGGCGCTGCCGGGCGCGAGG 187	NIH-MGC Project URL: http://mgc.ncl.nih.gov
245 AGGACAGTAGTGGCAATTCGAGAGCGTTTCCCTTTCAGCTGCGTAGAGCGTCCAGGAGGAGG 304	NIH-MGC Project URL: http://mgc.ncl.nih.gov
188 GGGAAAGGAGGCTGGAAATCCCAACACAGCTCCCTTTCAGAGACCCAGGCGTGCACAGACTTA 247	NIH-MGC Project URL: http://mgc.ncl.nih.gov
305 GAGAAAGGGGGCTTGAAGCACCAGCTGCCCTTTCAGAGCAGCAATGCTTCTCCAGAAATTGA 364	NIH-MGC Project URL: http://mgc.ncl.nih.gov
248 GCTGCTTAGAGAGAGGCGAGAAAGGCGCAATGGGAGAGACTCTGCTGCCGTGGCGGCACT 307	NIH-MGC Project URL: http://mgc.ncl.nih.gov


```

OY 721 TTCATGAGAGAGGGGTGAGGAGATG9GGGCGACGATGGAGTGGAGGGAGCGGAC 780
DB 72816 TTCGTGAGAAAGCTGTGTGAGAGAGACGGAGCGACGATGGCATGGAGGGAGCGGAC 72875
OY 781 GAGTTTTCGACGGGACTTTCGAGAGCTACGAGCGGTACCAACAGAGAGCGCTGCAG 840
DB 72876 GAGTTTTCGACGGGACTTTCGAGAGCTTACGAGCGGTACCAACAGAGCGCGTGCAG 72935
OY 841 AACATGAGAGAGAGAGAGCTCATCAAGAGTACTGGAAGTGGAAAGTGCCTTCGCGC 900
DB 72936 AACATGAGAGAGAGAGAGCTCATCAAGAGTACTGGAAGTGGAAAGTGCCTTCGCGC 72995
OY 901 ATGAGAGAGAGAGAGAGAGCGCTGCGCTGAGAGAGAGAGAGCGCTGGTGGCAGCAGCG 960
DB 72996 AAGGAGAGAGAGAGAGAGAGCGCTGCGCTGAGAGAGAGAGAGCGCTGGTGGCAGCAGCG 73055
OY 961 CGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 73056 CGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 73115
OY 1021 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 73116 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 73175

RESULT 10
AX331371 461 bp DNA linear PAT 09-JUN-2002
LOCUS Sequence 1880 from Patent WO0194629.
ACCESSION AX331371
VERSION AX331371.1 GI:18122005
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.,
Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL
Avalon Pharmaceuticals (US)
Patent: WO 0194629-A 1880 13-DEC-2001.
FEATURES
source
location/Qualifiers
1..461
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 105 a 134 c 138 g 80 t 4 others
ORIGIN
Query Match 32.9%; Score 355.2; DB 6; Length 461;
Best Local Similarity 95.7%; Pred. No. 2,1e-49;
Matches 374; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

```

```

OY 301 GCGCACTTCCCGCCCGGAGAACTGGAGACCGAGCCCGAGCGAGCTGCGCCAG 360
DB 372 -GCGACTTCCCGGCGGAGAAAGTGGAGCGGAGCCCGAGGCGGAGCTGTCAG 430
OY 361 CCTTGTATGACTCCGAGGAGAGTAAGTTGG 391
DB 431 CCTTGTATGACTCCGAGGAGAGTAAGTTGG 461

RESULT 11
BC025970 1238 bp mRNA linear PRI 22-MAR-2002
LOCUS Homo sapiens, similar to putative, clone MGC:39389 IMAGE:4559410,
DEFINITION mRNA, complete cds.
ACCESSION BC025970
VERSION BC025970.1 GI:19684066
KEYWORDS MGC.
SOURCE
ORGANISM
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 1238)
AUTHORS
Strasberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (18-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
Info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Seedi, Jacqueline
Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAL Plate: 42 Row: d Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
source
location/Qualifiers
1..1238
/organism="Homo sapiens"
/db_xref="locusID:124790"
/db_xref="taxon:9606"
/clone="MGC:39389 IMAGE:4559410"
/tissue_type="Eye, retinoblastoma"
/clone_lib="N1H-MGC_16"
/lab_host="DH10B-R"
/notes="vector: pORF7"
90..950
/codon_start=1
/product="similar to putative"
/protein_id="AAH25970.1"
/db_xref="GI:19684067"
/translaction="MMAFPNCTACNAESPVALLEAKTSGAPSPQTPPERHDSGLP
LTPRMESHEDDLAVAGLGNRSRPTOSPGGCSAVALARAKHRRSPKRRHW
RPLRLSNAEKOROSORASRYREEMFAKGVAVPYNTQFLMNRDPEPNLDV
PHGTHPSGSGSEAGSDSDGRARAGGORDFTSTYRFRHESIOSGSKELVVDYL
ELEKRLSQAEETRLQDLQCTGGGSCQVDELAAYVORLATEMORLROENQMRRE

```

GCRDEPCT*			
BASE COUNT	324 a	338 c	392 g
ORIGIN	184 t		
Query Match	16.2%; Score 174.8; DB 9; Length 1238;		
Best Local Similarity	60.5%; Pred. No. 1.6e-19;		
Matches 362; Conservative 0; Mismatches 197; Indels 39; Gaps 3;			
OY	441 GCTGGGAGAGAAAAACATAAGAGACGCCGCTCCAAAGAGAGCGGCGATTGGAAACCGTA	500	
DB	344 GCTGGCCCGGAGAAACACCGCTCGCGCCATCCAAAGCGAAAGGACTGGCCACCTTA	403	
OY	501 CTACAGCTGAACCTGGAGAGAGAAAAAGTTTCAGACAGAAACAGACCTTCGAGCTTC	560	
DB	404 CTTGAGACTGAGCTGGGCTGAGAAACAACCGGGATGAGAGGACCCAGAGGCGCTC	463	
OY	561 AAGATCCGAGCCGAGATGTTGGCAAGGCGCAGCGGCTGCGCCCTATAACACACGCA	620	
DB	464 CCGGGTCCGGAGAGATGTTGCGCAAGGCGACCGCTGGCCCCCTACAAACACACCGCA	523	
OY	621 GTTCCTATGATGATCATCAGACCGAGAGAGACCGGATCTCAAAACCGGCTGTACTCAA	680	
DB	524 GTTCTCATGATGATGACAGGACCGGAGAGCC-----CAACTTGAGATGTCGCCCAT	575	
OY	681 GCGGCGCCGCGCAAAATCCGACAGACAGCATGACACTTTCATGAGAAAGAGGGGTGA	740	
DB	576 GGGATCTCCACCCACAGTTCCAG-----TGGGGAGAG	607	
OY	741 GGAGATGGGGCGGCGCATGGAGTGGAGGGGACGCGAGCTTTCAGCGGGACTT	800	
DB	608 TGAGGCCGGGAGACAGTATGGGCGGCGCGGCGCACGCTGAGTTCCAGCGGAGGACTT	667	
OY	801 CTGGAGACGACGAGGGGTACACACGAGGAGGCTCAGAACATGACGAGCAGAGAGCT	860	
DB	668 CTCTGACCTTACGACAGCTTCCACACCGAGAGGCTTAGGGCCGCGCAAGAGGAGCT	727	
OY	861 CATCAGAGTACTCTGGAACGTGAGAGTGCCTCTCCGCGATGAGAGACGAGCAACCG	920	
DB	728 GGTCCGAGACTACTGGAGCTGAGAAAGCGGCTGTCCAGCGGAGGAGGAGACTAGAG	787	
OY	921 GCTCGGCTGAGAGCAAGCGGCTGGTGGCGA---CGACGCGGCTGTCCGGAGCTGGA	977	
DB	788 GCTGCACAGCTGAGAGCGCTGCACCGGCGACAGTCTTCGCCGAGGTGGAGAGCTGCG	847	
OY	978 GCTGAGCTGGAGCCGCGCGCGCGGAGAACCTCAGCTGTGTACCGCAGAGCAAGTG	1035	
DB	848 TGCCGAGGTCCAGAGGCTCCGGAGCCGAAACACAGCGGCTTGTGAGGAGAACAGATG	905	
RESULT 12			
AK056946	1330 bp	mRNA	linear PRI 01-AUG-2002
LOCUS			
DEFINITION	Homo sapiens cDNA FLJ32384 f1s, clone SKMUS1000104, weakly similar		
ACCESSION	AK056946		
VERSION	AK056946.1 GI:16552484		
KEYWORDS	oligo capping; f1s (full insert sequence).		
SOURCE	Homo sapiens skeletal muscle cDNA to mRNA, clone_lib:SKMUS1 clone:SKMUS1000104.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K., Arita,M., Muroshino,K., Youki,H., Hara,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Magatsune,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kanakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuko,Y., Nagai,K. and Isogai,T.		
TITLE	NEO human cDNA sequencing project		
JOURNAL	unpublished		

REFERENCE	2 (bases 1 to 1330)	
AUTHORS	Isogai,T., Otsuki,T. and Sugiyama,T.	
TITLE	Direct Submission	
JOURNAL	Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)	
COMMENT	NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.	
FEATURES	Location/Qualifiers	
Source	1..1330	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="SKMUS1000104"	
	/tissue_type="skeletal muscle"	
CDS	/clone_lib="SKMUS1"	
	/note="Cloning vector: pME18SFL3"	
	240..1100	
	/note="unnamed protein product"	
	/codon_start=1	
	/protein_id="BAB71319.1"	
	/db_xref="GI:16552485"	
	/translation="MMATPNOTACNAESPYALAEAKTSAPGSPQPPERHDSGSLP LTPMESHSEDEDLGAVGVGIMNSRPKUSPDCGSAEVLAKRRRPSKRRHM RPYLELSMAEQOQDEROSORASHREMEAKGPVAPYNTGTFILNDRPPEPNIDV PHGISHPGSGSEFAGSDGSRRAHGFORKDFSEYERFTELSIGRSKOEIVRDYL ELKRLSQAEEETFRLLQQLACTQGSQSRQVEELAAVORLRTENORLROENQMWNE GCRDEPCT"	
BASE COUNT	295 a	370 c 452 g 213 t
ORIGIN		
Query Match	16.2%; Score 174.8; DB 9; Length 1330;	
Best Local Similarity	60.5%; Pred. No. 1.6e-19;	
Matches 362; Conservative 0; Mismatches 197; Indels 39; Gaps 3;		
OY	441 GCTGGGAGAGAAAAACATAAGAGACGCCGCTCCAAAGAGAGCGGCGATTGGAAACCGTA	500
DB	494 GCTGGCCCGGAGAAACACCGCTCGCGCCATCCAAAGCGCACTGGCGCACCTTA	553
OY	501 CTAAAGCTGACCTGGAGAGAGAGAAAAAATTGAGAGAGAAACAGAGCTTCGAGCTTC	560
DB	554 CTTGAGCTGAGCTGGGCTGAGAAACAACCGGAGATGAGAGGACAGAGGCGCTC	613
OY	561 AAGATCCGAGCCGAGATGTTCCGCAAGGCGCAGCGGCTGCGCCCTATAACACACGCA	620
DB	614 CCGGGTCCGCAAGATGTTCCGCAAGGCGCACCGCTGCGCCCTACAAACACACCGCA	673
OY	621 GTTCCTATGATGATCATCAGACCGAGAGAGCGGATCTCAAAACCGGCTGTACTCAA	680
DB	674 GTTCTCATGATGATGACAGGACCGCGAGAGCC-----CAACTTGAGTGTGCCCAT	725
OY	681 GCGGCGCGCCGCAAAATCCAGACGACACCGAGATGAGAGCTTCTGGAAGAAGAGGGGTGA	740
DB	726 GGGATCTCTCCACCCAGGTTCCAG-----TGGGGAGAG	757
OY	741 GGAGATGGGGCGGCGCATGGAGTGGAGAGGAGCGGACGCGAGTTCGACGCGGACTT	800
DB	758 TGAGGCCGGGAGAGATGATGGGCGGCGGCGGCGGCGGCGGAGGAGGAGACTT	817
OY	801 CTGCGAGACGTAAGAGCGGTAACACACGAGAGAGCGCTTCAGAACATGAGCAAGAGAGCT	860
DB	818 CTCTGAGACTTACGAAGGCTTCCACACCGAGAGCGCTCAGGCGCGCAAGCAGAGAGCT	877
OY	861 CATCAAGAGTACTGGAACGTGAGAGAGTGCCTTCGCCCATGAGGAGACGAGAACCG	920
DB	878 GGTCCGAGACTTACGAGCTGGAGAGGCGGCTGTCTCCAGCGGAGGAGGAGTACGAG	937
OY	921 GCTGGCGCTGAGAGCAAGCGGCTGGGTGGCGA---CGACCGCGCTGTGGGAGGACTGGA	977

Db 938 GCTGAGAGCTGACAGCGCTGACCGGCGAGCAGTCTCCGCGCAGAGTGGAGAGCTGCG 997
 QY 978 GCTGAGAGCTGACAGCGCTGACCGGCGAGCAGTCTCCGCGCAGAGTGGAGAGCTGCG 1035
 Db 998 TGCCGAGGTCGACAGCGCTGACCGGCGAGCAGTCTCCGCGCAGAGTGGAGAGCTGCG 1055
 RESULT 13
 AC008105
 LOCUS AC008105 116035 bp DNA linear HTG 27-JUN-2002
 DEFINITION Homo sapiens chromosome 17 clone CTD-2020K17 map 17, *** SEQUENCING
 IN PROGRESS ***, 5 unordered pieces.
 AC008105
 AC008105.24 GI:21617753
 HTG: HTGS_PHASE1, HTGS_FULFILL, HTGS_ACTIVEFIN.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 116035)
 Birren, B., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 17, clone CTD-2020K17
 Unpublished
 2 (bases 1 to 116035)
 Birren, B., Linton, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
 Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., DeRellano, K., Depayre, E., Devon, K., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
 Hagos, B., Heaford, A., Horton, J., Howland, J., Jones, C., Kann, L.,
 Karas, A., Lehoczek, J., Lieu, C., Locke, K., Macdonald, P.,
 Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
 Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Teste, S., Toriunella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
 Direct Submission
 TITLE Submitted (22-JUL-1999) Whitehead Institute/MIT Center for Genome
 JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 3 (bases 1 to 116035)
 AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhaltier, B.,
 Camarato, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cooke, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
 Horton, L., Hulme, W., Hume, L., Johnson, R., Jones, C., Kamat, A.,
 Karas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Nord, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
 Teste, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zalnoun, J.,
 Zemek, L., Zimmer, A. and Zody, M.
 TITLE Direct Submission
 JOURNAL Submitted (27-JUN-2002) Whitehead Institute/MIT Center for Genome
 COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 27, 2002 this sequence version replaced g1:20148201.
 All repeats were identified using RepeatMasker:
 Smith, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 1890
 Center clone name: 2020_K_17

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 16830: contig of 16830 bp in length
 * 16831 16930: gap of 100 bp
 * 16931 29738: contig of 12808 bp in length
 * 29739 29838: gap of 100 bp
 * 29839 39151: contig of 9313 bp in length
 * 39152 39251: gap of 100 bp
 * 39252 45913: contig of 6662 bp in length
 * 45914 46013: gap of 100 bp
 * 46014 116035: contig of 70022 bp in length.
 *
 FEATURES
 source
 1..116035
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="CTD-2020K17"
 /clone_1lb="CTD1 Human BAC"
 BASE COUNT 26107 a 30142 c 30260 g 29051 t 475 others
 ORIGIN
 Query Match 16.2% Score 174.8; DB 2; Length 116035;
 Best Local Similarity 60.5% Pred. No. 8.7e-20;
 Matches 362; Conservative 0; Mismatches 197; Indels 39; Gaps 3;
 QY 441 GCTGGGGAAGAAAAACATAGAGAGCCCGTCCAGAGAGAGCGCATTTGAACCGTA 500
 Db 17959 GCTGGGCGGAGAAACACCGTGGCGGCGCATTCAGAAAGGCGACCTGCGACCTTA 18018
 QY 501 CTACAGCTGAACTGGGAGAGAGAGAAAGTTCCAGCAGAAAGAGCGCTTCGAGCTTC 560
 Db 18019 CTTGAGCTGACTGTGGCTGAGAAACACAGCGGATGAGAGCGACAGAGGCGCTTC 18078
 QY 561 AAGGATCCGAGCGGAGATGTCGCCAAGAGGCGCGTGGCGCCCTATACACAGCA 620
 Db 18079 CCGGGTCCCGGAGAGATTTGGCCAAAGCCAGCCCGCTTACACACAGCA 18138
 QY 621 GTTCTCATGATGATGATCAGCAGCAGAGAGCGGATCTCAAAACCGGCTTACTCAA 680
 Db 18139 GTTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 18190
 QY 681 GCGGGCGCGCGCGCAATCCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 740
 Db 18191 GGGATCTCCACCCAGCTTCAG-----TGGGAGAG 18222
 QY 741 GGAGATGGGGGAGCGATGGATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 800
 Db 18223 TGAGAGCGGGGAGAGATGATGGGCGGCGGAGCGGAGCGGAGGAGGAGGAGGAGGAG 18282
 QY 801 CTGCGAGCTAGCAGAGCGGTACCAACAGGAGAGCTTCAGAAACATGAGCAGCAGAGCT 860
 Db 18283 CTCTGAGACTTACGAACCTTCACACAGAGAGCTTCGAGGCGCCAGCAACAGAGAGCT 18342
 QY 861 CATCAAGAGCTACCTGAGAGCTGAGAGAGTCTCTCCGCAATGAGAGAGAGAGAGAGAG 920
 Db 18343 GGTGCGAGCTACTCTGAGAGCTGAGAGAGGCTGTGTCGAGGAGAGAGAGAGAGAG 18402
 QY 921 GCTGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 977
 Db 18403 GCTGACAGAGCTGACAGGCGTGCACCGGCGCAGCAGAGTCTGCGCGCAGTGGAGAGAGCTGCG 18462


```

QY      978 GCTGAGCTGACCGCTGCGCCGAGAACCTCCAGCTGCTGACGAGACGAACGTG 1035
      ||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | |
Db      18463 TGCGAGAGCTCCAGAGCGCTCGGAGACCGAAACACGCGCTTCGTCAGGACGAACGAGATG 18520

RESULT 14
AK096785/c
LOCUS      AK096785
DEFINITION Homo sapiens cDNA FLJ39466 fls, clone PROST2012353.
ACCESSION AK096785
VERSION    AK096785.1 GI:21756354
KEYWORDS   oligo capping; fls (full insert sequence).
SOURCE     Homo sapiens prostate cDNA to mRNA, clone_11d:PROST2
      clone:PROST2012353.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1
AUTHORS   Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizunuchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahara,K., Masuko,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2048)
TITLE      Isogai,T. and Yamamoto,J.
REFERENCE 1
AUTHORS   Isogai,T. and Yamamoto,J.
JOURNAL    Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
     source
         1..2048
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="PROST2012353"
         /tissue="prostate"
         /clone_lib="PROST2"
         /note="cloning vector: pME18SFL3"
BASE COUNT      383 a      655 c      558 g      452 t
ORIGIN
Query Match      16.0%; Score 173.2; DB 9; Length 2048;
Best Local Similarity 60.4%; Pred. No. 2.7e-19;
Matches 361; Conservative 0; Mismatches 198; Indels 39; Gaps 3;

QY      441 GCTGGGAGAAATAAATAGAGAGCGCCGTCACAGAAAGAGCGATTGGAACCGTA 500
      ||||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | |
Db      1650 GCTGGCGCCGGAACAAACCCGTGGCGGCATGAAAGCGCAAAAGGACACTGGCGACCTTA 1591

QY      501 CTACAAGCTGAACCTGGGAAGAGAAAGAAAGTTGACGAGAAACAGAGCCTTGAGACTTC 560
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1590 CCGTGAAGCTGAGCTGGGGCTGAGAAACACACGCGGAGTGAAGGACGACGAGCGGCTC 1531

QY      561 AAGGATCCGAGCCGGAATGTTCCGCAAGGCGCAGCGGCTGCGCCCTATTAACACCAAGCA 620
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1530 CCGGGTCCGGAAGAAATGTTCCGCAAGGCGCAGCGGCTGCGCCCTACACACCAACCA 1471

QY      621 GTTCCCTCATGATGATCAGACGACGAGGACCGGATCTCAAAACGGGCGCTGACTACCCA 680
      ||||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | |
Db      1470 GTTCTCGATGATGATGACAGGACCGCGGAGGAGCC-----CAACTTGATGTGGCCCAT 1419

QY      681 GCGGGCGCGCCCAATCCGACGACACGACGATGACGACTTATGTAAGAGGGGCTGA 740

```

```

Db      1418 GGGATCTCCACACCAGGTTTCAG-----TGGGAGAG 1387

QY      741 GAGAGATGGGGGACGACCATGATGGAGGGGACGCGACGAGTTTCTGACGGGAGCTT 800
      ||||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | |
Db      1386 TGAGGCCCGGGAGCATGATGAGCGCGGAGCGACGAGCGATGATGATCCAGCGGAAGGACTT 1327

QY      801 CTCGGAGACTACGAGCGGTACCAACGAGAGACCTTGCAACATATGACGAACGAGACT 860
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1326 CTCGAGACTTACGAGCGCTTCCACACGAGACCTTCAGAGGCGCGACGAAGGAGACT 1267

QY      861 CATCAAGAGTACCTGGAACTGGAAAGTCTCTGCGCATGAGAGACGAGCAACACG 920
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1266 GGTGCGAGACTACTCTGAGAGCTGAGAAAGCGCGCTGTGCGACGAGAGAGAGACTAGAG 1207

QY      921 GCTGCGGCTGGAGACGACGCGCTGGGTGGCGA---CGACGCGCGTGTGCGGAGCTGGA 977
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1206 GCTGACAGCACTCCAGCGCTCCGACCGCGCGAGCTGCTGCGCGCAGGTGAGAGAGCTGGC 1147

QY      978 GCTGAGAGCTGACCGGCTGCGCCGAGAACCTCCAGCTGCTGACGAGACGAACGTG 1035
      ||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | | ||| | |
Db      1146 TGCGAGAGTCCAGAGGCTCCGAGACCGAAACAGCGGCTTCTGTCAGAGAACGAGATG 1089

RESULT 15
LOCUS      AL662804
DEFINITION Mouse DNA sequence from clone Rp23-341C5 on chromosome 11, complete
sequence.
ACCESSION AL662804
VERSION    AL662804.17 GI:21621620
KEYWORDS   HMG
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 223442)
REFERENCE 1
AUTHORS   Blakey,S.
JOURNAL    Direct Submission
Submitted (26-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
on Jun 27, 2002 this sequence version replaced gi:21615602.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep Rp23-341C5 is
from the Rp21-23 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
FEATURES
     source
         1..223442

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="Rp23-341C5"
/clone_lib="RP23-23"
BASE COUNT 56058 a 54948 c 56347 g 56089 t
ORIGIN

Query Match 15.9% Score 172: DB 10: Length 223442;
Best Local Similarity 59.28; Pred. No. 2.3e-19;
Matches 354; Conservative 0; Mismatches 205; Indels 39; Gaps 2;

```
QY 441 GCTGGGGAAGAAAAACATACAGACGCCCGTCCAAAGAAAGACCGCATTTGGAACCGTA 500
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 3891 GCTGGCCCGAAAGAAAGCAACCCATAGGGCGGCATCGAAGCGCAAGCGCACTGGGCCCTTA 3950
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 501 CTACAAGCTGAACCTGGGAAGAGAAAAAGTTCAACGAGAAACAGAGCCTTCGAGCTTC 560
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 3951 CTTGAGAGCTGACCTGGCGCGAGCAACAGCGAGATGAGAGCGAAGCCAGAGGGGCTTC 4010
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 561 AAGATCCGAGCCGAGATGTTGCGCAAGGGCCAGCGGTCGGCCCTTAACACCACGCA 620
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4011 TCGGTCCTGAGAGAGATGTTGCCAAAGGTACGCCCTGGCGCCTTACACACCACTCA 4070
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 621 GTTCTCATGATGATCAAGACAGAGAGCGGATCTCAAAACCGGCTGTACTCCA 680
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4071 GTTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4130
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 681 GGGGGCCCCCGCAATCCGAGACGACCAAGCATGACACTCATGGAAGAGGGGTGA 740
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4131 CCACTCTGGCTCC-----GGTGGGAGANA 4154
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 741 GGAGATGGGGGACAGATGGGATGGAGGGGAGCGGACGAGATTCTGCAGCGGACTT 800
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4155 CGAAGCAGGGGACAGATGATGGGCAAGGCCGAGCCATGGGGATTTCAAGAGAGGACTT 4214
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 801 CTCGAGAGCTAGCAGCGGTACCAACGAGAGCCTGCAGAAACATGAGCAAGCAGAGCT 860
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4215 CTCTAGAGGCTATGAGCGGTACCACTGAGAGCCTTCAGGGCCGCGAGCAAGCAGAGCT 4274
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 861 CATCAAGGAGTACCTGAGAACTGAGAGAGTCTCGGCGCATGAGAGAGCAACACCG 920
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4275 GGTGCGAGACTACCTGAGATCTAGAGAGGGGTCTGTACAGGCTGAGCAAGAAACTCGGAG 4334
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 921 GCT---GCGGCTGAGAGACAGCGGCTGAGTGGCGAGCGCGGCTGTGCGGAGCTGGA 977
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4335 GCTCGGAGAGCTGACAGGGTTCCTCCAGCAGGCAACCTGTCAACAGGTGGAAGAGCTGGC 4394
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 978 GCTGAGCTGAGACGGCTGCGCGCGAGAACTCCAGCTGCTGACCGAGAGAGAACTG 1035
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4395 TGCTGAGGTGAGAGGCTCAGAACTGAAAAACAGCGGCTGCTCAGAGAAACGAGATG 4452
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: June 18, 2003, 01:05:33
Job time : 2958 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 01:11:07 ; Search time 949 Seconds
(Without alignments)
1669.991 Million cell updates/sec

Title: US-09-972-758a-1
Perfect score: 1080
Sequence: 1 atgacgcagaccatctctgtc.....ttccaagtgtgagactag 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues
Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PC7_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PCUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1080	100.0	1080	9 US-09-972-758-1	Sequence 1, Appli
2	1076.8	99.7	2199	9 US-09-745-763-33	Sequence 33, Appl
3	1076.8	99.7	3624	9 US-10-205-823-171	Sequence 171, App
4	368.8	34.1	495	9 US-09-918-995-403	Sequence 403, App
5	362.8	33.6	414	10 US-09-983-965-4895	Sequence 4895, App
6	355.2	32.9	461	9 US-09-954-531-813	Sequence 813, App
7	68.4	6.3	8705	9 US-10-291-230-14	Sequence 14, Appl
8	67.8	6.3	4270	9 US-09-954-531-1352	Sequence 1352, Ap
9	67.8	6.3	4270	10 US-09-880-107-2427	Sequence 2427, Ap
10	60.8	5.6	2108	10 US-09-962-832-225	Sequence 225, App
11	58.4	5.4	2313	9 US-09-854-133-157	Sequence 157, App
12	58.4	5.4	2314	10 US-09-738-973-157	Sequence 157, App
13	58.4	5.4	2314	9 US-09-764-868-12	Sequence 12, Appl
14	53.6	5.0	607	9 US-10-043-487-122	Sequence 122, App
15	53.6	5.0	1609	9 US-10-043-487-105	Sequence 105, App
16	52	4.8	225883	9 US-10-175-523-57	Sequence 57, Appl
17	50.2	4.6	1949	9 US-10-274-968-1	Sequence 1, Appl
18	50.2	4.6	1949	10 US-09-740-027-1	Sequence 1, Appl
19	49.6	4.6	506	9 US-10-060-036-222	Sequence 222, App

20	49.6	4.6	1327	9 US-10-043-487-168	Sequence 168, App
21	49.4	4.6	594	9 US-10-123-155-10	Sequence 10, Appl
22	49.2	4.6	14800	10 US-09-954-456-1601	Sequence 1601, Ap
23	49	4.5	3951	9 US-10-029-115-1	Sequence 1, Appl
24	49	4.5	4414	9 US-10-029-115-3	Sequence 3, Appl
25	48.2	4.5	401	9 US-09-918-995-37237	Sequence 37237, A
26	48.2	4.5	549	9 US-09-991-936-1712	Sequence 1712, Ap
27	47.4	4.4	1101	10 US-09-874-062-2	Sequence 2, Appl
28	47.4	4.4	1482	9 US-09-789-390-64	Sequence 64, Appl
29	47.4	4.4	1649	9 US-09-789-390-62	Sequence 62, Appl
30	47.4	4.4	1717	9 US-09-789-390-69	Sequence 69, Appl
31	47.4	4.4	1787	9 US-09-789-390-66	Sequence 66, Appl
32	47.4	4.4	3735	9 US-09-789-390-12	Sequence 12, Appl
33	47.4	4.4	3822	9 US-09-789-390-10	Sequence 10, Appl
34	47.4	4.4	3912	9 US-09-789-390-8	Sequence 8, Appl
35	47.4	4.4	3999	9 US-09-789-390-6	Sequence 6, Appl
36	47.4	4.4	4033	9 US-10-029-115-5	Sequence 5, Appl
37	47.4	4.4	4133	9 US-09-291-417-11	Sequence 11, Appl
38	47.2	4.4	520	9 US-10-184-644-332	Sequence 332, App
39	47.2	4.4	520	9 US-10-184-634-332	Sequence 332, App
40	47	4.4	3511	10 US-09-753-831-3	Sequence 3, Appl
41	47	4.4	4184	9 US-09-927-827-42	Sequence 42, Appl
42	46.8	4.3	971	9 US-10-102-806-61	Sequence 61, Appl
43	46.8	4.3	2411	10 US-09-823-356-23	Sequence 23, Appl
44	46.8	4.3	3259	12 US-10-001-843-71	Sequence 71, Appl
45	46.8	4.3	24707	9 US-10-274-968-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-972-758-1
Sequence 1, Application US/09972758
Patent No. US20020160497A1
GENERAL INFORMATION:
APPLICANT: Case Western Reserve University
APPLICANT: Montano, Monica
APPLICANT: Miltman, Bryan
TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth
FILE REFERENCE: 27708/04004
CURRENT APPLICATION NUMBER: US/09/972,758
CURRENT FILING DATE: 2001-10-05
PRIORITY APPLICATION NUMBER: US 60/238,187
PRIORITY FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 1080
TYPE: DNA
ORGANISM: Homo sapiens
US-09-972-758-1

Query Match 100.0%; Score 1080; DB 9; Length 1080;
Best Local Similarity 100.0%; Pred. No. 2e-278;
Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGCCGAGCCATTCTTGTGAGATATCAACACCCAGCTTAACCTAGCACTACAGCT 60
DB	1	ATGGCCGAGCCATTCTTGTGAGATATCAACACCCAGCTTAACCTAGCACTACAGCT 60
QY	61	GCCTGCTCTCTCCAGGAAGAGCTGACCCCTGACGCGCCCGGCGGAGAGCGGCTG 120
DB	61	GCCTGCTCTCTCCAGGAAGAGCTGACCCCTGACGCGCCCGGCGGAGAGCGGCTG 120
QY	121	CCGAGAGGAGCAAGTAGTGAGCAATGAGAGCGCTCCCGCAGTTGGGTGCGCTCGGGG 180
DB	121	CCGAGAGGAGCAAGTAGTGAGCAATGAGAGCGCTCCCGCAGTTGGGTGCGCTCGGGG 180
QY	181	CCGAGAGGAGCAAGTAGTGAGCAATGAGAGCGCTCCCGCAGTTGGGTGCGCTCGGGG 240
DB	181	CCGAGAGGAGCAAGTAGTGAGCAATGAGAGCGCTCCCGCAGTTGGGTGCGCTCGGGG 240

QY	241	GAATCTACTCTCCGTAGAGAGAGGGCGAGAAAGGGCCAGATGGGGAGACATCGTCGCGTGGC	300
Db	241	GAATCTACTCTCCGTAGAGAGAGGGCGAGAAAGGGCCAGATGGGGAGACATCGTCGTCGTC	300
QY	301	GGCGACTTCCCGCGCGCGGAGAAAGTGGAAACCGAGCGCCGAGGCCAGCTGCTGCGCCAG	360
Db	301	GGCGACTTCCCGCGCGCGGAGAAAGTGGAAACCGAGCGCCGAGGCCAGCTGCTGCGCCAG	360
QY	361	CCTTGTCATGACTCCGAGGCCAGTAAAGTTGGGGGCTGCTGCGCGAGGGGGCGAAGAGAG	420
Db	361	CCTTGTCATGACTCCGAGGCCAGTAAAGTTGGGGGCTGCTGCGCGAGGGGGCGAAGAGAG	420
QY	421	TGGGACAGCAGCAGAGACAGCTGGGGAGAAAAAAACATTAAGAGAGCGCCGTCCAAGAG	480
Db	421	TGGGACAGCAGCAGAGACAGCTGGGGAGAAAAAAACATTAAGAGAGCGCCGTCCAAGAG	480
QY	481	AAGGGGCACTTGGAAAAACGCTACTACAACTGTAACCTGGGAAGAGAAAAAAGTTGCGAG	540
Db	481	AAGGGGCACTTGGAAAAACGCTACTACAACTGTAACCTGGGAAGAGAAAAAAGTTGCGAG	540
QY	541	AAACGAGACCTTCGAGCTTCAAGAGTCCGAGCCGAGATGTCGCAAGGGCGAGCGGTC	600
Db	541	AAACGAGACCTTCGAGCTTCAAGAGTCCGAGCCGAGATGTCGCAAGGGCGAGCGGTC	600
QY	601	GGCGCCCTAATACACACCGCAGTTCTCTCATGATGATCAGACCCAGAGAGACCCGGATCTC	660
Db	601	GGCGCCCTAATACACACCGCAGTTCTCTCATGATGATCAGACCCAGAGAGACCCGGATCTC	660
QY	661	AAAACCGGCTCTGCTACTCCAGCGGGCGCGCCCAAAATCCGAGAGACACCGAGATGAGAG	720
Db	661	AAAACCGGCTCTGCTACTCCAGCGGGCGCGCCCAAAATCCGAGAGACACCGAGATGAGAG	720
QY	721	TTTCATGAGAAAGGGGGGTGAGAGAGATGGGGGCAACGATGGGATGGAGGGGAGCGGAC	780
Db	721	TTTCATGAGAAAGGGGGGTGAGAGAGATGGGGGCAACGATGGGATGGAGGGGAGCGGAC	780
QY	781	GAGTTTCGACGCGGGACTTCTCGGAGACGTACGAGCGCTACACACGAGAGAGCCTGAG	840
Db	781	GAGTTTCGACGCGGGACTTCTCGGAGACGTACGAGCGCTACACACGAGAGAGCCTGAG	840
QY	841	AACATGAGCAAGCAGAGAGCTCATCAAGAGATGACTGGAATCTGGAGAAATGCTCTCGCGC	900
Db	841	AACATGAGCAAGCAGAGAGCTCATCAAGAGATGACTGGAATCTGGAGAAATGCTCTCGCGC	900
QY	901	ATGAGAGAGCAGAAACAAACCGGCTCGCGTGGAGAGCAAGCGGCTGGGTGGCGAGACGCG	960
Db	901	ATGAGAGAGCAGAAACAAACCGGCTCGCGTGGAGAGCAAGCGGCTGGGTGGCGAGACGCG	960
QY	961	CGTGTGGCGGAGCTGAGAGCTGAGACTGGACCGGCTGGCGCGCGCGAGAAACCTCCAGCTGTG	1020
Db	961	CGTGTGGCGGAGCTGAGAGCTGAGACTGGACCGGCTGGCGCGCGCGAGAAACCTCCAGCTGTG	1020
QY	1021	ACCGAGAACGAACTGACACCGGCGAGAGAGCGAGCGCGCTTCCAAATTGGAGACTAG	1080
Db	1021	ACCGAGAACGAACTGACACCGGCGAGAGAGCGAGCGCGCTTCCAAATTGGAGACTAG	1080

RESULT 2

US-09-745-763-33

Sequence 33, Application US/09745763

Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John M.

Lavallee, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Merberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

ENCODING THEM

NUMBER OF SEQUENCES: 219

Query Match	Best Local Similarity	Matches 1078:	Conservative	Score 99.7%;	Score 1076.8;	DB 10;	Length 2199;
1	ATGCGCCGAGCCATTCTTCTCAGAAATATACACACAGCTCAAACTAGCACTGTACAGT	60					
707	ATGCGCGGAGCCATTCTTCTCAGAAATATACACACAGCTCAAACTAGCACTGTACAGT	766					
61	GCTGCTGCTGTCCAGAGAGAGCTGTAACCCCTGAGCGCCCCGAGCGCGAGAGAGCGGGTG	120					
767	GCTGCTGCTGTCCAGAGAGAGCTGTAACCCCTGAGCGCCCCGAGCGCGAGAGAGCGGGTG	826					
121	CCCGAGAGAGCACTAGTGTGCAATCCAGAGCGTTTCCCACTGTTGGTGTGCGCTCCGGGG	180					
827	CCCGAGAGAGCACTAGTGTGCAATCCAGAGCGTTTCCCACTGTTGGTGTGCGCTCCGGGG	886					
181	CCGGAGGGGGAGAGGAGCGCTGGAATCCCAACACCTCCCTTGACAGACCAGCCGCTGTCCA	240					
887	CCGGAGGGGGAGAGGAGCGCTGGAATCCCAACACCTCCCTTGACAGACCAGCCGCTGTCCA	946					
241	GAATCTAGCTCCTGAGAGAGGGGCGAGAAAGGGCGAGAAATGGGGAGCACTCGTCCGTGGC	300					
947	GAATCTAGCTCCTGAGAGAGGGGCGAGAAAGGGCGAGAAATGGGGAGCACTCGTCCGTGGC	1006					
301	GGGCACTTCCCGCGCGCGCGCAAGTGTGAACCCAGACGCCCGAGGGCGAGAGTCTGTGGCCAG	360					
1007	GGGCACTTCCCGCGCGCGCGCAAGTGTGAACCCAGACGCCCGAGGGCGAGAGTCTGTGGCCAG	1066					
361	CCTTGATCATGACTCCGAGAGCCAGTAAGTGTGGGGGCTCCTGCGCGAGGGGGGAGAGAGAG	420					
1067	CCTTGATCATGACTCCGAGAGCCAGTAAGTGTGGGGGCTCCTGCGCGAGGGGGGAGAGAGAG	1126					
421	TGGGAGACGAGCAGAGACAGCTTGGGGAGAAAAAATTAAGAGAGACGCCGCTCCAAGAG	480					
1127	TGGGAGACGAGCAGAGACAGCTTGGGGAGAAAAAATTAAGAGAGACGCCGCTCCAAGAG	1186					
481	AAGGGCATTTGGAAGAACCTGTAATCAAACTGAACTGGGGAAGAGAAAAATTGTGACAGAG	540					
1187	AAGGGCATTTGGAAGAACCTGTAATCAAACTGAACTGGGGAAGAGAAAAATTGTGACAGAG	1246					
541	AAAGAGAGCCCTTCAGGCTTCAAGGATCCGAGGCCAGATGTGTTCCGCAAGGGGCGACCGGCTC	600					

```
|||||
Db 1247 AACAGAGCCTTCGAGCTTCAGAGATCCGAGCCGAGATGTCGCCAAGGGCCAGCGCGTC 1306
QY 601 GCGCCCTATACACACGAGAGTTCCTCATGATGATCAGACAGGAGAGCGCGATCTC 660
Db 1307 GGGCCCTATMACACGACGAGTTCCTCATGATGATCAGACAGGAGAGCGCGATCTC 1366
QY 661 AAAACCGGCTGTACTCCCAAGCGGGGGCGCCCAATCCGAGACACCGAGCATGACGAC 720
Db 1367 AAAACCGGCTGTACTCCCAAGCGGGGGCGCCCAATCCGAGACACCGAGCATGACGAC 1426
QY 721 TTCATGGAAGAGGGGGTGTAGAGAGATGGGGGACGATGGATGGAGGGAGCGGACAC 780
Db 1427 TTCATGGAAGAGGGGGTGTAGAGAGATGGGGGACGATGGATGGAGGGAGCGGACAC 1486
QY 781 GAGTTTCTGACGCGGAGCTTCTCGAGAGCTACGAGCGGTACACACGAGACCTGCAG 840
Db 1487 GAGTTTCTGACGCGGAGCTTCTCGAGAGCTACGAGCGGTACACACGAGACCTGCAG 1546
QY 841 AACATGAGCAGCAGGAGCTCATCAAGGAGTACTGGAACCTGGAGAGAGTGGCTTGGCGC 900
Db 1547 AACATGAGCAGCAGGAGCTCATCAAGGAGTACTGGAACCTGGAGAGAGTGGCTTGGCGC 1606
QY 901 ATGAGAGAGAGAACACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTTGGCAGCAGCGC 960
Db 1607 ATGAGAGAGAGAACACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTTGGCAGCAGCGC 1666
QY 961 CGTGTGCGGAGCTGAGAGCTGAGCTGACCGCGCTGCGCGCAGACACTCCAGCTGCTG 1020
Db 1667 CGTGTGCGGAGCTGAGAGCTGAGCTGACCGCGCTGCGCGCAGACACTCCAGCTGCTG 1726
QY 1021 ACCGGAACGAACTGACCGGAGCAGGAGCAGGAGCGCGCTTCCAAAGTTTGGAGACTG 1080
Db 1727 ACCGGAACGAACTGACCGGAGCAGGAGCAGGAGCGCGCTTCCAAAGTTTGGAGACTG 1786

RESULT 3
US-10-205-823-171
: Sequence 171, Application US/10205823
: Publication No. US20030108963A1
: GENERAL INFORMATION:
: APPLICANT: Schlegel, Robert
: APPLICANT: Monahan, John E.
: APPLICANT: Endeje, Wilson O.
: APPLICANT: Gannavarapu, Manjula
: APPLICANT: Gorbacheva, Bella
: APPLICANT: Hoersch, Sebastian
: APPLICANT: Kamalkar, Shubhangi
: APPLICANT: Monsey, Angela M.
: APPLICANT: Glatz, Karen
: APPLICANT: Zhao, Xumei
: APPLICANT: Anderson, Dustin
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
: TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: FILE REFERENCE: MRI-044
: CURRENT APPLICATION NUMBER: US/10/205, 823
: PRIOR APPLICATION NUMBER: 60/307, 982
: PRIOR FILING DATE: 2001-07-25
: PRIOR APPLICATION NUMBER: 60/314, 356
: PRIOR FILING DATE: 2001-08-22
: PRIOR APPLICATION NUMBER: 60/325, 020
: PRIOR FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: 60/341, 746
: PRIOR FILING DATE: 2001-12-12
: PRIOR APPLICATION NUMBER: 60/362, 158
: PRIOR FILING DATE: 2002-03-05
: NUMBER OF SEQ ID NOS: 455
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 171
: LENGTH: 3624
: TYPE: DNA
```

```
: ORGANISM: Homo sapiens
US-10-205-823-171
Query Match 99.7%; Score 1076.8; DB 9; Length 3624;
Best Local Similarity 99.8%; Pred. No. 1.9e-277;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCGAGCCATTCCTGTGACAGATATCAACCCAGCCCTCAAACTAGCAACTGTACAGGT 60
Db 690 ATGGCCGAGCCATTCCTGTGACAGATATCAACCCAGCCCTCAAACTAGCAACTGTACAGGT 749
QY 61 GCTGCTGCTGTCCAGGAAGAGCTGAACCTTGAGCGCCGCCAGCGCGGAGAGCGGGTGT 120
Db 750 GCTGCTGCTGTCCAGGAAGAGCTGAACCTTGAGCGCCGCCAGCGCGGAGAGCGGGTGT 809
QY 121 CCCGAGAGAGACAGTACGTGGCAATCGAGAGCGTTCCCAATGGGTGGCTCCGGGG 180
Db 810 CCCGAGAGAGACAGTACGTGGCAATCGAGAGCGTTCCCAATGGGTGGCTCCGGGG 869
QY 181 CCGGAGGGGGAGGAGGAGCTTGAATCCCAACCACTCCCTTGACAGCCAGGCTGTGCA 240
Db 870 CCGGAGGGGGAGGAGGAGCTTGAATCCCAACCACTCCCTTGACAGCCAGGCTGTGCA 929
QY 241 GAATCTAGCTGCTGAGAGAGGGCGAGAAAGGCCAATGGGGAGCACTCTCCGCTGGC 300
Db 930 GAATCTAGCTGCTGAGAGAGGGCGAGAAAGGCCAATGGGGAGCACTCTCCGCTGGC 989
QY 301 GCGCACTTCCGCGCGCGGAGAGATGGACCGACGCCCGAGCGGAGCTGCTCCGCCAG 360
Db 990 GCGCACTTCCGCGCGCGGAGAGATGGACCGACGCCCGAGCGGAGCTGCTCCGCCAG 1049
QY 361 CCTGTGATGATCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGGCGAAGAGAG 420
Db 1050 CCTGTGATGATCCGAGGCGCAGTAAGTTGGGGGCTCTGCGCGAGGGGGCGAAGAGAG 1109
QY 421 TGGGGACACAGCAGAGAGAGCTGGGGAAGAAAAACATAGAAGCGCCGTCGAAGAG 480
Db 1110 TGGGGACACAGCAGAGAGAGCTGGGGAAGAAAAACATAGAAGCGCCGTCGAAGAG 1169
QY 481 AAGCGCATTTGGAACCCGTACTACAACTGAACTGGGAAGACAAAGAAAGTTCCAGCAG 540
Db 1170 AAGCGCATTTGGAACCCGTACTACAACTGAACTGGGAAGAAAGAAAGTTCCAGCAG 1229
QY 541 AAACAGAGCCTTCGACCTTCAAGATCCGAGCGGAGATGTTGCCAAGGGCCAGCGGTC 600
Db 1230 AAACAGAGCCTTCGACCTTCAAGATCCGAGCGGAGATGTTGCCAAGGGCCAGCGGTC 1289
QY 601 GCGCCCTATMACACGACGAGTTCTCATGATGATCAGACAGCAGAGAGAGCGGATCTC 660
Db 1290 GCGCCCTATMACACGACGAGTTCTCATGATGATCAGACAGCAGAGAGAGCGGATCTC 1349
QY 661 AAAACCGGCTGTACTCCCAAGCGGGGGCGCCCAATCCGAGACACACGAGCATGACGAC 720
Db 1350 AAAACCGGCTGTACTCCCAAGCGGGGGCGCCCAATCCGAGACACACGAGCATGACGAC 1409
QY 721 TTCATGGAAGAGGGGGTGTAGAGAGATGGGGGACGATGGATGGAGGGAGCGGACAC 780
Db 1410 TTCATGGAAGAGGGGGTGTAGAGAGATGGGGGACGATGGATGGAGGGAGCGGACAC 1469
QY 781 GAGTTTCTGACGCGGAGCTTCTCGAGAGCTACGAGCGGTACACACGAGAGCTGCAG 840
Db 1470 GAGTTTCTGACGCGGAGCTTCTCGAGAGCTACGAGCGGTACACACGAGAGAGCTGCAG 1529
QY 841 AACATGAGCAGCAGGAGCTCATCAAGGAGTACTGGAACCTGGAGAGAGTGGCTTGGCGC 900
Db 1530 AACATGAGCAGCAGGAGCTCATCAAGGAGTACTGGAACCTGGAGAGAGTGGCTTGGCGC 1589
QY 901 ATGAGAGAGAGAAACACCGGCTGCGGCTGAGAGAGACAGCGGCTGGTGGCAGCAGCGC 960
Db 1590 ATGAGAGAGAGAAACACCGGCTGCGGCTGAGAGAGACAGCGGCTGGTGGCAGCAGCGC 1649
QY 961 CGTGTGCGGAGCTGAGAGCTGAGAGCTGACCGGCTGCGCGCGAGAACCTCCAGCTGCTG 1020
Db 1020 CGTGTGCGGAGCTGAGAGCTGAGAGCTGACCGGCTGCGCGCGAGAACCTCCAGCTGCTG 1020
```

Db 1550 CGTGTGGGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCGAGAACCTTCAGCTGCTG 1709
QY 1021 ACCGAGAACGAAGTGCACCGGCGACGAGGAGCGCGCTTTTCCAGTTTGAGACTAG 1080
Db 1710 ACCGAGAACGAAGTGCACCGGCGACGAGGAGCGCGCTTTTCCAGTTTGAGACTAG 1769

RESULT 4
US-09-918-995-403/c
Sequence 403, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 403
LENGTH: 495
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(495)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-403

Query Match 34.1%; Score 368.8; DB 9; Length 495;
Best Local Similarity 99.5%; Pred. No. 7.1e-89;
Matches 370; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 709 AGCATATGACGACTTCATGGAAGAGGGGTGAGAGATGGGGCAGCGATGGATGGGA 768
Db 495 AGCATATGACGACTTCATGGAAGAGGGGTGAGAGATGGGGCAGCGATGGATGGGA 436
QY 769 GGGAGCGGACGAGTTTCTGACGCGGACTTCTCGAGAGCTACGAGCGGTACACACG 828
Db 435 GGGAGCGGACGAGTTTCTGACGCGGACTTCTCGAGAGCTACGAGCGGTACACACG 376
QY 829 GAGAGCTGCGAAGATGAGCAAGCAGAGCTCATCAAGAGTACTGGAACCTTGAGAG 888
Db 375 GAGAGCTGCGAAGATGAGCAAGCAGAGCTCATCAAGAGTACTGGAACCTTGAGAG 316
QY 889 TGCTCTCGCGCATGAGAGAGAGCAACCGGCTGCGGCTGAGAGCAACCGGCTGGT 948
Db 315 TGCTCTCGCGCATGAGAGAGAGCAACCGGCTGCGGCTGAGAGCAACCGGCTGGT 256
QY 949 GGGAGCGGACGAGTTTCTGACGCGGACTTCTCGAGAGCTGACCGGCTGCGCGGAGAAC 1008
Db 255 GGGAGCGGACGAGTTTCTGACGCGGACTTCTCGAGAGCTGACCGGCTGCGCGGAGAAC 196
QY 1009 CTCGAGCTGCTGACCGAGAACGAACTGACCGGCGAGAGGAGCGCGCTTTCCAG 1068
Db 195 CTCGAGCTGCTGACCGAGAACGAACTGACCGGCGAGAGGAGCGCGCTTTCCAG 136
QY 1069 TTTGGAGACTAG 1080
Db 135 TTTGGAGACTAG 124

RESULT 5
US-09-983-965-4895
Sequence 4895, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengping
APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 4895
LENGTH: 414
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 24-LIB34-008-Q1-E1-F7
US-09-983-965-4895

Query Match 33.6%; Score 362.8; DB 10; Length 414;
Best Local Similarity 92.3%; Pred. No. 2.7e-87;
Matches 382; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 462 GAGACGCGCCCTCCAGAGAAAGCGGCAATGGAAACGCTACTACAGCTGAAGTGGAGAGA 521
Db 1 GAGACGCGCCCTCCAGAGAAAGCGGCTTGGAAACGCTACTACAGCTGAAGTGGAGAGA 60
QY 522 GAAGAAAAGTTGACGAGAAACAGAGCCTTCGAGTTCAAGATCCGACCGAGATGT 581
Db 61 GAAGAAAAGTTGACGAGAAACAGAGCCTTCGAGTTCAAGATCCGACCGAGATGT 120
QY 582 CGCCAGAGGCGAGCGGCTGCGCCCTTAAACACGAGAGTTCTCATGTGATGACGA 641
Db 121 CGCCAGAGGCGAGCGGCTGCGCCCTTAAACACGAGAGTTCTCATGTGATGACGA 180
QY 642 CCAGAGAGCGGATGCAAAACCGGCTGATCTCAAGGCGGCGCGCAAAATCGA 701
Db 181 CCAGAGAGCGGATGCAAAACCGGCTGATCTCAAGGCGGCGCGCAAAATCGA 240
QY 702 CGACACGAGGATGACGATTCATGGAAGAGGGGTGAGAGAGTGGGGCAGCGATGG 761
Db 241 CGACACGAGGATGACGATTCATGGAAGAGGGGTGAGAGAGTGGGGCAGCGATGG 300
QY 762 GATGGAGAGGAGCGGACGCGAGTTTCTGACGCGGACTTCTCGAGACGTAAGCGGTA 821
Db 301 GATGGAGAGGAGCGGACGCGAGTTTCTGACGCGGACTTCTCGAGACGTAAGCGGTA 360
QY 822 CCACAGGAGAGCGCTGCAAGACATGACGACGAGAGCTCATCAAGAGTACCT 875
Db 361 CCACAGGAGAGCGCTGCAAGACATGACGACGAGAGCTCATCAAGAGTACCT 414

RESULT 6
US-09-954-531-813
Sequence 813, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
FILE REFERENCE: 689290-77
CURRENT APPLICATION NUMBER: US/09/954,531
CURRENT FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/60/233,133
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,009
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,034
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,509
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US/60/234,567
PRIOR FILING DATE: 2000-09-22

```

: NUMBER OF SEQ ID NOS: 1392
:
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 813
: LENGTH: 461
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1)..(461)
: OTHER INFORMATION: n-a,t,g or c
: OS-09-954-531-813

```

Query Match	32.9%	Score 355.2	DB 9	Length 461
Best Local Similarity	95.7%	Pred. No. 3e-85		
Matches 374	Conservative	0	Mismatches 16	Indels 1
				Gaps 1

QY	1	ATGGCCGAGCCATTCTTGTACAAATATCAACACGAGCCCTCAAACTGCAACTGTACAGGT	60
Db	72	ATGGCCGAGCCATTCTTGTACAAATATCAACACGAGCCCTCAAACTGCAACTGTACAGGT	131
QY	61	GCTGCTGCTGTCCAGGAAGAGCTGAACCTTGAGCGCCGCCACGAGCGCGGAGAGACGCGGTG	120
Db	132	GCTGCTGCTGTCCAGGAAGAGCTGAACCTTGAGCGCCGCCACGAGCGCGGAGAGCGGTG	191
QY	121	CCCGAGCAGGACAGTAGTGGCAATCGAGAGCGTTGCCCAAGTTGGGTGGCGTCCGGGG	180
Db	192	CCCGAGCAGGACAGTAGTGGCAATCGAGAGCGTTGCCCAAGTTGGGTGGCGTCCGGGG	251
QY	181	CCGGAGGGGGAAGGGAGCGCTGGAAATCCCAACCAACGCTCCCTGCAGACCCGAGCGCTGTCCA	240
Db	252	CCGGAGGGGGAAGGGAGCGCTGGAAATCCCAACCAACGCTCCCTGCAGACCCGAGCGCTGTCCA	311
QY	241	GAATTCAGCTGCTCCTAGAGAGGGGCGCAAGAGGCGCAGAAATGGGGACCACTGTCGCTGGC	300
Db	312	GAATTCAGCTGCTCCTAGAGAGGGGCGCAAGAGGCGCAGAAATGGGGACCACTGTCGCTGGC	371
QY	301	GGCGACTTCCCGCCCGCCGAGAGTGAACCGACGCGCCGAGGCGGAGGCTGTCGCCCCAG	360
Db	372	-GCGCACTTCCCGCCCGCCGAGAGTGAACCGACGCGCCGAGGCGGAGGCTGTCGCCCCAG	430
QY	361	CCTTGTCATGACTCCGAGGCGCAGTAGTTGG	391
Db	431	CCTTGTCATGACTCCGAGGCGCAGTAGTTGG	461

```

1 RESULT 7
2 US-10-291-230--14/C
3 ? Sequence 14, Application US/10291230
4 ? Publication NO. US20030108939A1
5 ? GENERAL INFORMATION:
6 ?
7 ? APPLICANT: Ruffner, Duane E.
8 ? APPLICANT: Pierce, Michael L.
9 ? APPLICANT: Chen, Zhidong
10 ?
11 ? TITLE OF INVENTION: Directed Antisense Libraries
12 ? FILE REFERENCE: T6678.US.A
13 ? CURRENT APPLICATION NUMBER: US/10/291,230
14 ? CURRENT FILING DATE: 2002-11-07
15 ? PRIOR APPLICATION NUMBER: US 09/647,344
16 ? PRIOR FILING DATE: 2000-12-04
17 ? PRIOR APPLICATION NUMBER: PCT/US99/06742
18 ? PRIOR FILING DATE: 1999-03-28
19 ? PRIOR APPLICATION NUMBER: US 60/79,792
20 ? PRIOR FILING DATE: 1998-03-28
21 ? PRIOR APPLICATION NUMBER: US 60/107,504
22 ? PRIOR FILING DATE: 1998-11-06
23 ? NUMBER OF SEQ ID NOS: 50
24 ? SOFTWARE: PatentIn version 3.1
25 ?
26 ? SEQ ID NO 14
27 ?
28 ? LENGTH: 8705
29 ? TYPE: DNA
30 ? ORGANISM: Artificial Sequence
31 ? FEATURE:
32 ? OTHER INFORMATION: pshuttle

```

US-10-291-230-14

Query Match	6.38;	Score 68.4;	DB 9;	Length 8705;
Best Local Similarity	44.08;	Pred. No. 1.6e-08;		
Matches 288;	Conservative	0;	Mismatches 366;	Indels 0;
				Gaps 0;

OY	406	GGGGGCCAAGAGGAGTGTTGGGACACAGAACACAGACAGCAGCTGGGGGAAAGAAAAACATTAAGAGA	465
Db	7382	GGGCGAGACGACGGAGGGAGGGGCGAGACACAGAGAGGGGCGACGAGGGGCGACAGAGGGGCAC	7923
OY	466	CGCCCTCTCCAAGAAAGACGGCATTTGGGAAACCGTACTACAAAGCTGAATCTGGGAAGAGAG	525
Db	7922	GAGGGGCGACGAGGACGAGAGAGCGGCAGAGAGAGAGAGGGGCGCAGAGGGGCGACGAGAGGGG	7863
OY	526	AAAAAGTTCCAGCAGAAACAGAGCCCTTCGAGCTTCAAGGATCCGAGCCGAGATGTTTGCC	585
Db	7862	CAGGAGCAGGAGGAGGGGGCAGGAGCAGAGAGAGGGGCGAGAGGGGCGAGGACAGAGAGAG	7803
OY	586	AAGGGCCAGCCGGTGGCGCCCTTATACACCAGCGAGTTCCATGATGATACAGCAGC	645
Db	7802	GGGCGAGGAGGGGGCAGGAGGGGCGAGGAGCAGAGAGAGGGGCGCAGAGGACAGAGAGGGGCAG	7743
OY	646	GAGGAGCCGGATCTCAAAACCGGCGCTGTACTCCAAGCGGGCCCGCCCAATCCGACGAC	705
Db	7742	GAGGGGCGACGACGAGAGAGGGCGCAGAGGGGCGAGAGGGGCGAGGAGCGACAGAGAGAGGGG	7683
OY	706	ACCAAGCATGTCGCACTTCATGGAAGAGGGGGGTGAGGAGATGGGGGCGACCATGATGGGATG	765
Db	7682	CAGGACGACGGGGGGGCGAGAGGGGCGAGAGGGGCGAGAGCGCAGGAGGGGGCGAGACACAGAG	7623
OY	766	GGAGGGGACGCGACAGCGAGTTTCTTCACCGGGACTTCTCGAGACGTAACGACGGCTTACAC	825
Db	7622	GAGGGGCGAGGAGGGGCGACGAGGGCGCAGGAGGCGCAGAGGGGCGACGAGAGGGGCGAGGAG	7563
OY	826	ACGGAAGACCTTCGCAAAACATGACCAACGACGAGAGCTCTCAAGGAGTACTCTTGGAACCTGGAG	885
Db	7562	CAGGAGGGGGCGAGCGACGAGAGGGGCGAGAGGGGCGAGGAGCGAGAGGGGCGAGAGGGGCGAG	7503
OY	886	AAGTGTGCTCTGGCCCATGAGAGGAGGACCAACCGGGTGGCGGCTGTGAGAGCAACAGCGCTG	945
Db	7502	GAGCAGGAGGGGGCGACGAGGGGCGAGGACAGGAGGAGGGGCGAGAGGGGCGAGGACAGGAGAG	7443
OY	946	GGTGGCCACGACGCGCGCTGTGTGCGGGAGCTGGAGCTGGAGCTGAGCTGACCGAGCTGCGCGCCAG	1005
Db	7442	GAGGGGCGAGGGGGCGACGAGACACAGAGGGGCGAGAGGGGCGAGGAGCGACGAGAGCGCAGAGAG	7383
OY	1006	AACCTCCAGCTGCTGACCGAGAGAACGAACGTACCGCGGACGACGAGAGCCAGCGCCG	1059
Db	7382	GGGCGAGGACGAGAGGGGCGAGAGGGGCGAGGAGGAGGAGGAGGAGGGGCGAGGAGAG	7129

```

: RESULT 8
: US-09-954-531-1352
: Sequence 1352, Application US/09954531
: Patent No. US20020165180A1
: GENERAL INFORMATION:
: APPLICANT: Weaver, Zoe
: TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
: TITLE OF INVENTION: Gene Sets
: FILE REFERENCE: 689290-77
: CURRENT APPLICATION NUMBER: US/09/954, 531
: CURRENT FILING DATE: 2002-05-02
: PRIOR APPLICATION NUMBER: US/60/233,133
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: US/60/234, 009
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234, 034
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234, 509
: PRIOR FILING DATE: 2000-09-22
: PRIOR APPLICATION NUMBER: US/60/234, 567
: PRIOR FILING DATE: 2000-09-22
: NUMBER OF SEQ ID NOS: 1392

```

SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1352
; LENGTH: 4270
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-1352

Query Match 6.3%; Score 67.8; DB 9; Length 4270;
Best Local Similarity 49.9%; Pred. No. 2e-08;
Matches 171; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 712 GATGACGACTTATGGAAGAGGGGTGAGAGATGGGGGACGAGATGGATGGAGGG 771
DB 439 GAGTACGACTTCAAGATGCGACAGCGCTGAGAGAGCTGCGGACGACAGCAGCA 498
QY 772 GAGCGACGAGCTTCTGAGCGGGGACTTCTGGAGACCTAGACGCGGACCAACAG 831
DB 499 GTGGCGCTCTACAGCTGAGAGCTGGAGCAGACCTACAGAGCCAAAGTGGACG 558
QY 832 AGCCGTGACAGACATGAGCAAGCAGAGCTCATCAAGAGTACCTGGAAGTGC 891
DB 559 CTGAGCTCTGACCAAGACAGAGCGGCGGCTCGGAGAGAGCTGAAGAGGCGC 618
QY 892 CTCTGGCGCATGAGAGAGCAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTC 951
DB 619 CGCATGCGCGCTGAGTCCCTCAGCTACCAAGCTCCCGGCTCCAGAGAGCGCCAGTGC 678
QY 952 GACGAGCGCGCTGTCGCGAGCTGAGAGCTGAGAGCTGAGAGCGGCTGCGGCGAGAACCTC 1011
DB 679 GCTGAAGATCGCATTCGGAGAGCTGGAGAGGCGCCATGCGCGGAGCGGAGCAAGTTCCGG 738
QY 1012 CAGCTGCTGACGAGAGCAAGCACTGCAACCGCGAGCAGAGCGAG 1054
DB 739 AAGATGCTGAGCGCCAGAGAGCAGAGATGATGAGAGATGCGGG 781

RESULT 9

US-09-880-107-2427
; Sequence 2427, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2427
; LENGTH: 4270
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M94362
US-09-880-107-2427

Query Match 6.3%; Score 67.8; DB 10; Length 4270;
Best Local Similarity 49.9%; Pred. No. 2e-08;
Matches 171; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 712 GATGACGACTTATGGAAGAGGGGTGAGAGATGGGGGACGAGATGGATGGAGGG 771
DB 439 GAGTACGACTTCAAGATGCGACAGCGCTGAGAGAGCTGCGGACGACAGCAGCA 498
QY 772 GAGCGACGAGCTTCTGAGCGGGGACTTCTGGAGAGCTAGAGAGCGGATGACAGCGAG 831

DB 499 GTGGCGCTCTACAGCTGAGAGCTGAGAGCACTACCAAGGCCAGCTGGACAGCCCAAG 558
QY 832 AGCCGTGACAGACATGAGCAAGCAGAGGCTCATCAAGAGTACTTGAAGTGC 891
DB 559 CTGAGCTCTGACCAAGACAGAGCGGCGGCTGAGAGAGCTGCGGAGAGACTGAAGAGGCGC 618
QY 892 CTCTGGCGCATGAGAGAGCAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTC 951
DB 619 CGCATGCGCGCTGAGTCCCTCAGCTACCAAGCTCTCCGCGCTCCAGAGAGCGCCAGTGC 678
QY 952 GACGAGCGCGCTGTCGCGAGCTGAGAGCTGAGAGCTGAGAGCGGCTGCGGCGAGAACCTC 1011
DB 679 GCTGAAGATCGCATTCGGAGAGCTGGAGAGGCGCCATGCGCGGAGCGGAGCAAGTTCCGG 738
QY 1012 CAGCTGCTGACGAGAGCAAGCACTGCAACCGCGAGCAGAGCGAG 1054
DB 739 AAGATGCTGAGCGCCAGAGAGCAGAGATGAGAGATGCGGG 781

RESULT 10

US-09-962-832-225
; Sequence 225, Application US/09962832
; Patent No. US20020110821A1
; GENERAL INFORMATION:
; APPLICANT: Ederer, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-74
; CURRENT APPLICATION NUMBER: US/09/962,832
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,077
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,280
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 259
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 225
; LENGTH: 2108
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-832-225

Query Match 5.6%; Score 60.8; DB 10; Length 2108;
Best Local Similarity 51.1%; Pred. No. 1.3e-06;
Matches 143; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 789 GCAGCGGAGCTCTGCGAGAGCTAGAGCGGCTGACACAGCGAGAGCTGCAGAACATGAG 848
DB 782 GCAGCTGAGAGCTCTGAGAGCAGCAGAGAGAGAGCTGAAGCAGCTGGAGCAGCAGAGAGG 841
QY 849 CAGCAGAGAGCTCATCAAGAGAGTACCTGGAAGTGGAGAGTGCCTCTGCGCATGGAGGA 908
DB 842 GCAGCTGAGAGGTCGCCAGAGAGCAGATGGGAGAGCTGAAGTACCTGGAAGCAGAGAGAGG 901
QY 909 CGAGAACAAACCGGCTGCGGCTGGAGAGCAGCGGCTGGGTCGACAGAGCGGCTGTCG 968
DB 902 GCAGCTGAGAGCACTCTGATCAGCAGAGAGCAAGCAGCAAGCTCCAGAGCAGCAGATGGG 961
QY 969 GGAGCTGAGAGCTGAGAGCTGAGAGCGGCTGCGGCGGAGAACCTTCAGCTGAGCAGAGAA 1028
DB 962 GCAGCTGAGAGCACTCTGAGAGCAGCAGAGAGAGAGGAGCAGCTTAAGCATCTGGAGCAGCAGAGAGG 1021
QY 1029 CGAACTGACCGGCGAGCAGAGAGCGAGCGCGCTTTCAGAG 1068
DB 1022 GCAACTGGAGCAGAGCTGAGAGAGCAGAGAGGCGGAGCTGAAG 1061

RESULT 11

US-09-854-133-157
; Sequence 157, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.


```

; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 157
; LENGTH: 2313
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-854-133-157

```

```

Query Match      5.4%; Score 58.4; DB 9; Length 2313;
Best Local Similarity 52.5%; Pred. No. 5.6e-06;
Matches 128; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```

```

QY      817 CGGTACACACGAGAGCCTGCGAATGAGACAGAGAGCTCATCAAGAGTACTG 876
Db      968 CGGCTCAGAGCCGAGGGAAGACGCTCTACACAGAGCTGAAGCAACGCGCGAG 1027
QY      877 GAACGTGAGAAAGTGCCTCTGCGCATGAGAGAGCAACACCGCTGCGGTGAGAGC 936
Db      1028 CACGGGAGAGAGGAGGAGCGCGCCGCGCCCAAGAGAGAGAGCTGCTGCGGCTGCGAG 1087
QY      937 AACGGCTGGTGGCGACGAGCGCGCTGCGGAGCTGAGCTGAGACCGGCTG 996
Db      1088 CAGCTCAGAGAGAGAGGAGCGAGAGCTGACAGAGCTGAGAGCTGCTGCGAGGCGGAG 1147
QY      997 CGGCGGAGAGACCTGAGCTGACGAGCAACAGAGAGCTGAGCGGAGAGAGCG 1056
Db      1148 CGGAGAGCGAGGCTGCTGCGAGAGAGAGAGAGAGAGCGCGCCAGACCGCGGAG 1207
QY      1057 CCGC 1060
Db      1208 CTGC 1211

```

```

RESULT 12
US-09-738-973-157
; Sequence 157, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Joseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 157
; LENGTH: 2313
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-738-973-157

```

```

Query Match      5.4%; Score 58.4; DB 10; Length 2313;

```

```

Best Local Similarity 52.5%; Pred. No. 5.6e-06;
Matches 128; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```

```

QY      817 CGGTACACACGAGAGCCTGCGAATGAGACAGAGAGCTCATCAAGAGTACTG 876
Db      968 CGGCTCAGAGCCGAGGGAAGACGCTCTACACAGAGCTGAAGCAACGCGCGAG 1027
QY      877 GAACGTGAGAAAGTGCCTCTGCGCATGAGAGAGCAACACCGCTGCGGTGAGAGC 936
Db      1028 CACGGGAGAGAGGAGGAGCGCGCCGCGCCCAAGAGAGAGAGCTGCTGCGGCTGCGAG 1087
QY      937 AACGGCTGGTGGCGACGAGCGCGCTGCGGAGCTGAGAGCTGAGACCGGCTG 996
Db      1088 CACCTCAGAGAGAGAGAGGAGCGGAGAGCTGACAGAGCTGAGAGCTGCTGCGAGGCGGAG 1147
QY      997 CGGCGGAGAGACCTGAGCTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1056
Db      1148 CGGAGAGCGAGGAGGCTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1207
QY      1057 CCGC 1060
Db      1208 CTGC 1211

```

```

RESULT 13
US-09-764-868-12
; Sequence 12, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 2314
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-12

```

```

Query Match      5.4%; Score 58.4; DB 9; Length 2314;
Best Local Similarity 52.5%; Pred. No. 5.6e-06;
Matches 128; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```

```

QY      817 CGGTACACACGAGAGCCTGCGAATGAGACAGAGAGCTCATCAAGAGTACTG 876
Db      976 CGGCTCAGAGCCGAGGGAAGAGCTCTACACAGAGCTGAAGCAACGCGCGAG 1035
QY      877 GAACGTGAGAAAGTGCCTCTGCGCATGAGAGAGCAACACCGCTGCGGTGAGAGC 936
Db      1036 CACGGGAGAGAGGAGGAGCGCGCCGCGCCCAAGAGAGAGAGAGAGAGAGAGAG 1095
QY      937 AACGGCTGGTGGCGACGAGCGCGCTGCGGAGAGCTGAGAGCTGAGACCGGCTG 996
Db      1096 CACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1155
QY      997 CGGCGGAGAGACCTGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1056
Db      1156 CGGAGAGCGAGGAGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1215
QY      1057 CCGC 1060
Db      1216 CTGC 1219

```

```

RESULT 14
US-10-043-487-122
; Sequence 122, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:

```

APPLICANT: HYBRIGENICS
APPLICANT: PIERRE, LEGRAIN
TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptide
TITLE OF INVENTION: mammalian polypeptides
FILE REFERENCE: B4778A
CURRENT APPLICATION NUMBER: US/10/043,487
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/261,130
PRIOR FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 561
SOFTWARE: PatentIn version 3.1
SEQ ID NO 122
LENGTH: 607
TYPE: DNA
ORGANISM: Shigella Flexneri
US-10-043-487-122

Query Match 5.0%; Score 53.6; DB 9; Length 607;
Best Local Similarity 52.2%; Pred. No. 8.1e-05;
Matches 119; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 841 AACATGACGACGAGCTCATCAAGAGTACCTGGAAGTGGAGAGTCCCTCTGCGC 900
DB 230 AAGATGAGGACAGCATTAAGAGAGGCGGAAGTACGAGATTAAACCGTGGCAGATC 289
QY 901 ATGAGACGACAGAACACCGGCTGCGCTGAGAGCAAGCGGCTGGGTGGCGAGCGC 960
DB 290 ACTGAGTCACTCACTCCATCCGTCGTGAGAGAGGCCAGAGCAGCGGCATGACCGT 349
QY 961 CGTGTGGGAGCTGAGCTGAGCTGAGACCGGCTGCGCGAGAACTCCAGCTGCTG 1020
DB 350 CTTGTGCTGGTGAGAGAGGCTCTGCAACAGCTGGCAGAAAGAGAGGCCCAAGCTGCTG 409
QY 1021 ACCGAGAACAACTGCACGCGCAGCAGAGCGCGCTTTCCAG 1068
DB 410 GCCCAGCTGGCCAGAGTGTCAAGAGCAGCGGCGAGGCTCCCCAG 457

RESULT 15
US-10-043-487-105
Sequence 105, Application US/10043487
Publication No. US20030055220A1
GENERAL INFORMATION:
APPLICANT: HYBRIGENICS
APPLICANT: PIERRE, LEGRAIN
TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptide
TITLE OF INVENTION: mammalian polypeptides
FILE REFERENCE: B4778A
CURRENT APPLICATION NUMBER: US/10/043,487
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/261,130
PRIOR FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 561
SOFTWARE: PatentIn version 3.1
SEQ ID NO 105
LENGTH: 1609
TYPE: DNA
ORGANISM: Shigella Flexneri
US-10-043-487-105

Query Match 5.0%; Score 53.6; DB 9; Length 1609;
Best Local Similarity 52.2%; Pred. No. 9.9e-05;
Matches 119; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 841 AACATGACGACGAGCTCATCAAGAGTACCTGGAAGTGGAGAGTCCCTCTGCGC 900
DB 1232 AAGATGAGGACAGCATTAAGAGAGGCGGAAGTACGAGATTAAACCGTGGCAGATC 1291
QY 901 ATGAGACGACAGAACACCGGCTGCGCTGAGAGCAAGCGGCTGGGTGGCGAGCGC 960
DB 1292 ACTGAGTCACTCACTCCATCCGTCGTGAGAGAGGCCAGAGCAGCGGCATGACCGT 1351
QY 961 CGTGTGGGAGCTGAGCTGAGCTGAGACCGGCTGCGCGAGAACTCCAGCTGCTG 1020

DB 1352 CTTGTGCTGGGACGAGAGGCTCTGCAACAGCTGGCAGAGAGAGGCCCAAGCTGCTG 1411
QY 1021 ACCGAGAACAACTGCACCGGCGCAGCAGAGCGCGCTTTCCAG 1068
DB 1412 GCCCAGCTGGCCAGAGTGTCAAGAGCAGCGGCGAGGCTCCCCAG 1459

Search completed: June 18, 2003, 05:28:17
Job time: 964 secs

C	10	82.8	7.7	567	22	AAH12885	Human cDNA clone
	11	68.4	6.3	799	19	AAV56831	Nucleotide sequence
	12	68.4	6.3	1926	21	AAAS0254	Epsstein Barr Virus
	13	68.4	6.3	1926	22	AAF82902	EBV tethering protein
	14	68.4	6.3	2560	21	AAAT7542	Nucleotide sequence
	15	68.4	6.3	2560	24	AAI64275	Epsstein-Barr Virus
C	16	68.4	6.3	5452	20	AAAX90923	Anti-sense strand
	17	68.4	6.3	8705	20	AAZ23778	Vector pShuttle DN
	18	68.4	6.3	9600	19	AAV21683	Vector plasmid pCVC
	19	68.4	6.3	10380	20	AAZ22248	Nucleotide sequence
	20	68.4	6.3	10596	14	AAO51731	Plasmid pcISEBON f
	21	68.4	6.3	10596	17	AAAT40348	Plasmid pcISEBON f
	22	68.4	6.3	10596	20	AAAI5650	Nucleotide sequence
C	23	68.4	6.3	16080	21	AAAS5953	DNA clone pCEK-C1
	24	67.8	6.3	4270	24	AABN95930	Gene #2428 used to
	25	67.8	6.3	4270	24	ABL64082	Breast cancer rela
	26	67.8	6.3	5250	22	AAKS2371	Human polynucleot
	27	67.8	6.3	5254	22	AAKS5355	Human polynucleot
	28	67.8	6.1	9551	20	AAK22301	cDNA encoding a hu
	29	65.2	6.0	1925	20	AAAX90824	cDNA encoding a hu
	30	65	6.0	795	19	AAV55830	Epslein Barr Virus
	31	63.4	5.9	3489	21	AAAS0290	FLGA insert stabl
	32	63.4	5.9	3489	22	AAAF82901	Kaposi's sarcoma-e
	33	63.4	5.9	3489	24	AABA93487	Nucleotide sequenc
C	34	63.4	5.9	32207	20	AAV73805	Kaposi's sarcoma-e
	35	63.4	5.9	137507	19	AAVI9941	KSHV LUR DNA (nuc
C	36	60.8	5.6	2108	24	ABJ67774	KSHV long unique c
	37	60.2	5.6	2744	16	AAQ98470	Oesophagus cancer
	38	60.2	5.6	5120	22	AAAC84677	MSP1-containing f
	39	58.4	5.4	1896	21	AAAC28684	DNA sequence of hu
	40	58.4	5.4	2313	20	AAZ07197	JEST open reading
	41	58.4	5.4	2313	21	AACT9151	Human lung tumour
	42	58.4	5.4	2313	23	AAAD23226	Human lung tumour
	43	58.4	5.4	2314	22	AAAS26977	cDNA encoding nove
	44	58.4	5.4	2802	22	AAKS2420	Human polynucleot
	45	58.4	5.4	2853	21	AAA28683	DNA encoding JEST

ALIGNMENTS

RESULT 1							
ABN84013	ID	ABN84013 standard; cDNA; 1080 BP.					
XX	AC	ABN84013;					
XX	DT	10-SEP-2002 (first entry)					
XX	DE	Human oestrogen downregulated gene EDG1 coding sequence.					
XX	ED	EDG1: oestrogen downregulated gene; tumour suppressor; human;					
KW	breast cancer; prostate cancer; testicular cancer; ovarian cancer;						
KM	uterine cancer; colon cancer; chromosome 17q; gene therapy; gene;						
XX	ss.						
XX	OS	Homo sapiens.					
XX	Key	Location/Qualifiers					
FT	CDS	1..1080					
FT		/*tag= a					
FT		/product= "EDG1"					
FT		/transl_except= "(pos:460..462,aa:Arg)"					
FT		/transl_except= "(pos:511..513,aa:Thr)"					
FN		WO200228879-A1.					
PD		11-APR-2002.					
PF		05-OCT-2001; 2001WO-US31300.					
PR		05-OCT-2000; 2000US-238187P.					

PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
XX Montano M, Wiltman B;
XX WPI, 2002-519107/55.
DR P-PSDB; ABB76495.
XX
XX Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,
PT useful for the prevention, diagnosis and treatment of e.g. breast
PT cancer, testicular cancer, prostate cancer, uterine cancer, cervical
PT cancer, ovarian cancer and colon cancer -
XX
XX Claim 1; Fig 1A-B; 52pp; English.
PS
XX
XX The present sequence is the coding sequence for human oestrogen
CC downregulated gene 1 (EDG1), a tumour suppressor gene that is
CC downregulated by oestrogen in mammary epithelial cells. The gene
CC was identified by yeast two-hybrid screenings for oestrogen
CC receptor-interacting proteins in breast epithelial cells. It
CC was localised to chromosome arm 17q. EDG1 mRNA expression is
CC prevalent in normal mammary epithelial cells and in other human
CC hormone-responsive tissues such as the ovary, prostate and testis.
CC Expression is low in breast cancer epithelial cells. Oestradiol,
CC which induces breast cancer cell growth, has an inhibitory effect
CC on EDG1 mRNA expression in breast cancer cells. Hexamethylene
CC bis-acetamide, an inducer of differentiation and apoptosis,
CC upregulates EDG1 mRNA expression in breast cancer cells. The
CC invention provides EDG1 polynucleotides and polypeptides. In a
CC claimed method, a test sample from an individual suspected of
CC having, or known to have breast, testicular, prostate, uterine,
CC cervical, ovarian or colon cancer is assayed for EDG1 transcript
CC using a polynucleotide that is complementary to the present
CC sequence or by RT-PCR using a primer derived from the present
CC sequence. A decrease in the level of transcript compared to the
CC level in a test sample indicates that the test sample contains or
CC was derived from cancerous cells antibody. A claimed method for
CC decreasing the proliferation of breast, prostate, testicular,
CC ovarian, uterine, cervical or colon cancer cells involves increasing
CC EDG1 protein activity in the cells, either by contacting the cells
CC with EDG1 protein or its fragment or functional equivalent, or with
CC a nucleic acid encoding EDG1 protein, its fragment or functional
CC equivalent.
XX
XX Sequence 1080 BP; 265 A; 296 C; 376 G; 143 T; 0 other:
SQ
Query Match 100.0%; Score 1080; DB 24; Length 1080;
Best Local Similarity 100.0%; Pred. No. 8.7e-214;
Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 CCTTGTCATGACTCCGAGGCGAGTAGTAAGTTGGGGGCTCCTGCCGAGGGGGGAGAGAG 420
DB 361 CCTTGTCATGACTCCGAGGCGAGTAGTAAGTTGGGGGCTCCTGCCGAGGGGGGAGAGAG 420
QY 421 TGGGACAGCAGCAGACAGCTGGGGGAAGAAAACTAAGACAGCCCTCCAAAGAG 480
DB 421 TGGGACAGCAGCAGACAGCTGGGGGAAGAAAACTAAGACAGCCCTCCAAAGAG 480
QY 481 AAGCGGCAATTGGAAACCGTACTACAGCTGAACCTGGGAAGAGAAAAAGTTGCAGAG 540
DB 481 AAGCGGCAATTGGAAACCGTACTACAGCTGAACCTGGGAAGAGAAAAAGTTGCAGAG 540
QY 541 AAACAGAGCCCTTGAGCTTCAAGATCCAGCCGAGATGTTGCCAAGGCCAGCGGCTC 600
DB 541 AAACAGAGCCCTTGAGCTTCAAGATCCAGCCGAGATGTTGCCAAGGCCAGCGGCTC 600
QY 601 GCGCCCTTAAACACAGCAGTCTCTCATGATGATCAAGCCAGAGAGAGCCGATCTC 660
DB 601 GCGCCCTTAAACACAGCAGTCTCTCATGATGATCAAGCCAGAGAGAGCCGATCTC 660
QY 661 AAACCGGCTTGTACTCCAAAGCGGCGCCCGCAATCCGACGACACCCAGGATGACGAC 720
DB 661 AAACCGGCTTGTACTCCAAAGCGGCGCCCGCAATCCGACGACACCCAGGATGACGAC 720
QY 721 TTCAATGAGAAAGGGGGTGAGAGAGATGGGGCAGCTGGAGTGGAGGGAGCGGAC 780
DB 721 TTCAATGAGAAAGGGGGTGAGAGAGATGGGGCAGCTGGAGTGGAGGGAGCGGAC 780
QY 781 GAGTTTGCAGCGGAGCTTTCGAGAGAGTACGAGCGGTACACAGCAGAGAGCTGAG 840
DB 781 GAGTTTGCAGCGGAGCTTTCGAGAGAGTACGAGCGGTACACAGCAGAGAGCTGAG 840
QY 841 AACATGACCAAGCAGAGCTCATCAAGAGTAGTACCTGGAATCGAGAAAGTGCCTCGGC 900
DB 841 AACATGACCAAGCAGAGCTCATCAAGAGTAGTACCTGGAATCGAGAAAGTGCCTCGGC 900
QY 901 ATGAGAGACAGAAACACCGGCTGCGCTGGAGAGCAAGCGGCTGGTGCGACAGCGC 960
DB 901 ATGAGAGACAGAAACACCGGCTGCGCTGGAGAGCAAGCGGCTGGTGCGACAGCGC 960
QY 961 CGTGGCGGAGCTGAGCTGAGCTGAGACCGGCTGCGCGCGCGAGAACCTCAGCTGCTG 1020
DB 961 CGTGGCGGAGCTGAGCTGAGCTGAGACCGGCTGCGCGCGCGAGAACCTCAGCTGCTG 1020
QY 1021 ACCGAGAAAGAACTGACACCGGAGCAGAGAGCGCGCTTTCAGATTGGAGACTAG 1080
DB 1021 ACCGAGAAAGAACTGACACCGGAGCAGAGAGAGCGCGCTTTCAGATTGGAGACTAG 1080

RESULT 2
AAV82778
ID AAV82778 standard; cDNA; 2199 BP.
XX
XX AAV82778;
AC
XX 25-FEB-1999 (first entry)
DT
XX
XX
DE Clone bp783_3 isolated from human foetal kidney cDNA library.
XX
XX
OS Secreted protein; nutritional activity; immune stimulating; vaccine;
XX suppressing activity; haematopoiesis regulating activity;
XX tissue growth activity; activin; inhibin activity; chemotaxis;
KW chemokine activity; haemostasis; thrombolytic activity; receptor;
KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW tumour inhibition; gene therapy; ds.
XX
XX Homo sapiens.
OS
XX
XX W09842739-A2.
PN
XX
XX 01-OCT-1998.
PD
XX
XX 20-MAR-1998; 98WO-0505653.
PF

XX 19-MAR-1998: 98US-0044466.
PR 21-MAR-1997: 97US-0822167.
XX
PA (GEMV) GENETICS INST INC.
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Werberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
XX MPI: 1998-609890/51.
DR P-PDB: AAM85455.
XX
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, placenta or adult
PT pineal gland cDNA libraries.
XX
PS Claim 1: Page 66-67: 113pp: English.
XX
CC The present sequence encodes a secreted protein. The polynucleotide and
CC secreted protein are predicted to have biological activities which would
CC make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is given.
CC Suggested activities include nutritional activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity (no data is
CC given in the specification to support these activities). The
CC polynucleotide is also stated to be useful for gene therapy.
XX
SO Sequence 2199 BP: 552 A; 511 C; 674 G; 462 T; 0 other;

Query Match 99.7%: Score 1076.8; DB 19; Length 2199;
Best Local Similarity 99.8%: Pred. No. 4.5e-213;
Matches 1078: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCGAGCATTCTTGTGAGATATCAACACCGCTCAAACTAGCACTGTACAGGT 60
DB 707 ATGGCCGAGCATTCTTGTGAGATATCAACACCGCTCAAACTAGCACTGTACAGGT 766
QY 61 GCTGCTGCTGTCAGAGAGAGCTGAACCTTGAGCGCCCCCAGCGCGGAGAGCGGGTG 120
DB 767 GCTGCTGCTGTCAGAGAGAGCTGAACCTTGAGCGCCCCCAGCGCGGAGAGCGGGTG 826
QY 121 CCGGAGAGAGAGAGAGAGCTGCAATGAGAGCGTTCCCGCAGTTGGTGCCTCGGGG 180
DB 827 CCGGAGAGAGAGAGAGAGCTGCAATGAGAGCGTTCCCGCAGTTGGTGCCTCGGGG 886
QY 181 CCGGAGAGAGAGAGAGAGCTGCAATGAGAGCGTTCCCGCAGTTGGTGCCTCGGGG 240
DB 887 CCGGAGAGAGAGAGAGAGCTGCAATGAGAGCGTTCCCGCAGTTGGTGCCTCGGGG 946
QY 241 GAATCTAGCTGCTGAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 947 GAATCTAGCTGCTGAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1006
QY 301 GGGGAGCTTCCCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 1007 GGGGAGCTTCCCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1066
QY 361 CTTTGTGATGACTTCGAGAGCGAGTAAAGTTGGGGGCTCTGCGGAGGAGAGAGAGAG 420
DB 1067 CTTTGTGATGACTTCGAGAGCGAGTAAAGTTGGGGGCTCTGCGGAGGAGAGAGAG 1126
QY 421 TGGGAGACAGCAGAGAGAGAGCTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 1127 TGGGAGACAGCAGAGAGAGCTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186
QY 481 AAGCGGATTGGAAACCGTACTACAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 1187 AAGCGGATTGGAAACCGTACTACAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1246

QY 541 AAACAGAGCCTTGAGCTTCAGAGATCCGAGCCGAGATGTTCCGCAAGGCGCAGCGGTC 600
DB 1247 AAACAGAGCCTTGAGAGTTCAAGAGATCCGAGCCGAGATGTTCCGCAAGGCGCAGCGGTC 1306
QY 601 GCGCCCTATTAACACACAGCAGGTTCTTCATGATGATCAGCAGCAGAGAGCGGATCTC 660
DB 1307 GCGCCCTATTAACACACAGCAGGTTCTTCATGATGATCAGCAGCAGAGAGCGGATCTC 1366
QY 661 AAAACCGGCTGTACTTCAAGCGCGCGCGCCCAATCCGACAGACACACACCATGAGAGAC 720
DB 1367 AAAACCGGCTGTACTTCAAGCGCGCGCGCCCAATCCGACAGACACACCATGAGAGAC 1426
QY 721 TTGATGAAGAGAGAGAGAGAGAGATGGGAGCAGCATGAGATGGAGAGAGAGAGAGAG 780
DB 1427 TTGATGAAGAGAGAGAGAGATGGGAGCAGCATGAGATGGAGAGAGAGAGAGAGAGAG 1486
QY 781 GAGTTTCTGAGCGGAGCTTCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 1487 GAGTTTCTGAGCGGAGCTTCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1546
QY 841 AACATGACAGCAGAGAGCTCATCAGAGATACCTGGAAGAGAGAGAGAGAGAGAGAGAG 900
DB 1547 AACATGACAGCAGAGAGCTCATCAGAGATACCTGGAAGAGAGAGAGAGAGAGAGAG 1606
QY 901 ATGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 1607 ATGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1666
QY 961 CGTGTGCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 1667 CGTGTGCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1726
QY 1021 ACCGAGAGCAACTGCAACCGGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1727 ACCGAGAGCAACTGCAACCGGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1786

RESULT 3
AB092015
ID AB092015 standard; cDNA: 2199 BP.
XX
AC AB092015;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 12.
XX
KW Human; cytosolic; antirheumatic; antiarthritic; vulnery; analgesic;
KW antinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antidiabetic; antidiabetic; antidiabetic; antidiabetic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002065394-A1.
XX
PD 30-MAY-2002.
XX
PF 22-DEC-2000; 2000US-0745763.
XX
PR 18-MAR-1998; 98US-0040963.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.

PA (MERB/) MERBERG, D.
PA (TREAC/) TREACY, M.
PA (SPAU/) SPAULDING, V.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;
XX
DR WPI: 2002-582343/62.
XX
DR P-PSDB: ABP61799.

PT Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis

Claim 37; Page 111-112; 284pp; English.

The invention relates to human secreted or transmembrane protein (I), their fragments and is encoded by specific complementary deoxyribonucleic acid (cDNA) inserts (II), where the protein is substantially free from other mammalian proteins. (I) are useful for preventing, treating or ameliorating a medical condition, especially immunological treatment or prevention of tumours. (I) exhibits activity relating to angiogenesis, cytokine, cell proliferation, cell differentiation, anti-inflammatory, stem cell growth factor activity and activin or inhibin-related activities. (I) can be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. (I) induces the proliferation of neural cells and regeneration of nerve and brain tissue and is useful for the treatment of central and peripheral nervous system diseases and neuropathies, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis. (I) is involved in chemotactic or chemokinetic activity, regulation of haematopoiesis and is useful for treating myeloid or lymphoid cell disorders, platelet disorders such as thrombocytopenia and for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue growth and in tissue repair, healing of burns, incisions, ulcers, for treating osteoporosis, osteoarthritis, bone degenerative disorders or periodontal disease. (I) is also useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, various immune deficiencies and disorders including severe combined immunodeficiency (SCID), bacterial or fungal infections, autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic reactions and conditions such as asthma or other respiratory problems. (II) is useful to express recombinant protein, as markers for tissues in which the corresponding protein is preferentially expressed and in gene therapy. The present sequence is that of a polynucleotide of the invention.

Sequence 2199 BP; 552 A; 511 C; 674 G; 462 T; 0 other;

Query Match	99.78;	Score 1076.8;	DB 24;	Length 2199;
-------------	--------	---------------	--------	--------------

```
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

OY	1	ATGGCGAGCATTCTTGTGCAGAAATTCACACACCGCCCAACCTAGCAACTGTACAGCT	60
Db	707	ATGGCGAGCATTCTTGTGCAGAAATTCACACACCGCCCAACCTAGCAACTGTACAGCT	766
OY	61	GCTGCTCTGTCCAGGAAGAGCTGAACCCCTGAGCGGCCCCACGAGCGGAGAGCGGGTG	120
Db	767	GCTGCTCTGTCCAGGAAGAGCTGAACCCCTGAGCGGCCCCACGAGCGGAGAGCGGGTG	822
OY	121	CCGAGAGGAGCAGTAGTGTGCATATGAGAGAGTTTCCCAAGTTGGTGGGCGTCCGGGG	180
Db	827	CCGAGAGGAGCAGTAGTGTGTGCATATGAGAGAGTTTCCCAAGTTGGTGGGCGTCCGGGG	885
OY	181	CCGAGAGGAGGAGAGGAGCCTGGCAATCCCAACCACTCCCTTGCAGAACCCAGGCTGTCCA	240
Db	887	CCGAGAGGAGGAGAGGAGCCTGGCAATCCCAACCACTCCCTTGCAGAACCCAGGCTGTCCA	944
OY	241	GAATCTAGCTGCTTGAGAGAGGCGGAGAAAGGCGCAGAAATGGGAGGATCTCGCTGGCG	300

Db	947	GAATCTAGTCTCTGAGAGAGGGGAGAAAGGGCCAGAAATGGGGACGACTGTCGGCTGGC	1006
QY	301	GGCGCACTTCCCGCCCGCGGCAAGTGGAAACCCAGCCCGCCAGGCGCGAETCTCTGCCAG	360
Db	1007	GGCACAATTCCCGCGCGCGGAGAAATGGAAACCGACGCCCGGAGCGCGAETCTCTGCCAG	1066
QY	361	CCTTGTCATGACTCCGAGGCGCAGTAAGTTGGGGGCTCCTGGCGAGGAGGGCGAAGAGAG	420
Db	1067	CCTTGTCATGACTCCGAGGCGCAGTAAGTTGGGGGCTCCTGGCGAGGAGGGCGAAGAGAG	1126
QY	421	TGGGAGACGACGACGAGACAGCTGGGGGAAGAAAAACATTAAAGACGCGCTCCAGAG	480
Db	1127	TGGGAGACGACGACGAGACAGCTGGGGGAAGAAAAACATTAAAGACGCGCTCCAGAG	1186
QY	481	AAGGGGCACTTGGAAACCGTACTTACAAAGCTGAACTGGGAAGAAAGAAAAATTTCAGCAG	540
Db	1187	AAGGGGCACTTGGAAACCGTACTTACAAAGCTGAACTGGGAAGAAAGAAAAATTTCAGCAG	1246
QY	541	AAACAGAGCCCTTGAGCTTCAAGAGTCCGAGCCCGAGATGTTCCGCCAAGGGGCGACCGGTC	600
Db	1247	AAACAGAGCCCTTGAGCTTCAAGAGTCCGAGCCCGAGATGTTCCGCCAAGGGGCGACCGGTC	1306
QY	601	GGCGCCCTTAACACGACGCAATTCCTCTATGATATACGACACGAGGAGGAGCGGAGTCTC	660
Db	1307	GGCGCCCTTAACACGACGCAATTCCTCTATGATATACGACACGAGGAGGAGCGGAGTCTC	1366
QY	661	AAAAACGGCCTGTACTCTCCAAACCGGGCGCGCCGCCAAATCCGACGACACACGCGATGACGAC	720
Db	1367	AAAAACGGCCTGTACTCTCCAAACCGGGCGCGCCGCCAAATCCGACGACACACGCGATGACGAC	1426
QY	721	TTCATGGAAGAAAGGGGTGTAGAGAGATGGGGGCAACGATGGGATGGAGGGGAGAGCGGAC	780
Db	1427	TTCATGGAAGAAAGGGGTGTAGAGAGATGGGGGCAACGATGGGATGGAGGGGAGAGCGGAC	1486
QY	781	GAGATTTTCGACGCGGAGCTTCTCGGAGACGTACAGAGCGGTACCAACAGCGAAGGCTCGAG	840
Db	1487	GAGATTTTCGACGCGGAGCTTCTCGGAGACGTACAGAGCGGTACCAACAGCGAAGGCTCGAG	1546
QY	841	AACATGAGCAAGCAGGAGCTCATCAAGAGTACTTGGAACTGGAGAAAGTGCCTCTCGGC	900
Db	1547	AACATGAGCAAGCAGGAGCTCATCAAGAGTACTTGGAACTGGAGAAAGTGCCTCTCGGC	1606
QY	901	ATGGAGGACGAGAAACAACCGGCTCGCGCTGGAGAGCAAGCGGCTGGGTTGGGAGCAGACGCG	960
Db	1607	ATGGAGGACGAGAAACAACCGGCTCGCGCTGGAGAGCAAGCGGCTGGGTTGGGAGCAGACGCG	1666
QY	961	CGTGTGCGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGCGGAACCTTCACACTGCTG	1020
Db	1667	CGTGTGCGGAGCTGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCCGCGGAACCTTCACACTGCTG	1726
QY	1021	ACCGGAACGAACTGCAACCGGCGACGAGAGCGGAGCGCGCTTTCCTCAAGTTTGGAGACTAG	1080
Db	1727	ACCGGAACGAACTGCAACCGGCGACGAGAGCGGAGCGCGCTTTCCTCAAGTTTGGAGACTAG	1786
RESULT 4			
AAH18098			
ID	AAH18098	standard; cdna: 2178 BP.	
XX	AAH18098;		
AC			
XX			
DT	26-JUN-2001	(first entry)	
XX			
DE	Human cdna sequence SEQ ID NO:17952.		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
XX			
OS	Homo sapiens.		
XX			
FN	EP1074617-A2.		
XX			
PD	07-FEB-2001.		
XX			

RESULT 4
AAH18098
ID AAH18098 standard; cDNA; 2178 BP

DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA sequence SEQ ID NO:17952

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens

PN EP1074617-A2

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17952; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2178 BP; 525 A; 513 C; 678 G; 462 T; 0 other:
Query Match 99.6%; Score 1075.2; DB 22; length 2178;
Best Local Similarity 99.7%; Pred. No. 9,6e-213;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 1016 GCGGACTTCCCGCCGCGGAGAGAGACCGAGCCCGAGCGGAGCTGCTGCCAG 1075
QY CTTGTGATGACTCCGAGGCCAGTAAATTGGGGGCTCTGCCGAGGGGGCGAGAGAG 420
DB 1076 CTTGTATGACTCCGAGGCCAGTAAATTGGGGGCTCTGCCGAGGGGGCGAGAGAG 1135
QY 421 TGGGAGACAGCAGACAGACAGCTGGGGGAAAGAAACATTAAGAGACGCCCTCCAAAG 480
DB 1136 TGGGAGACAGCAGACAGACAGCTGGGGGAAAGAAACATTAAGAGACGCCCTCCAAAG 1195
QY 481 AAGCGCATTTGGAACCTGACTACAACTGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAG 540
DB 1196 AAGCGCATTTGGAACCTGACTACAACTGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAG 1255
QY 541 AAACGAGCCTTGAGCTTCAAGATCCGAGCGAGATGTTGCCAAGGGCCAGCGGCTC 600
DB 1256 AAACGAGCCTTGAGCTTCAAGATCCGAGCGAGATGTTGCCAAGGGCCAGCGGCTC 1315
QY 601 GCGCCCTTAAACACCGAGCAGTCTCTCATGATGATGATCAAGACAGAGAGAGAGAGAGAG 660
DB 1316 GCGCCCTTAAACACCGAGCAGTCTCTCATGATGATGATCAAGACAGAGAGAGAGAGAGAG 1375
QY 661 AAACCGGCTGTACTCCAGCGGCGCGCCGCAANTCCGAGCAGCAGCAGGATGAGCAG 720
DB 1376 AAACCGGCTGTACTCCAGCGGCGCGCCGCAANTCCGAGCAGCAGGATGAGCAG 1435
QY 721 TTCAATGAGAGAGGGGTGAGAGATGGGGGACCGATGGAGTGGAGGGAGCGGACG 780
DB 1436 TTCAATGAGAGAGGGGTGAGAGATGGGGGACCGATGGAGTGGAGGGAGCGGACG 1495
QY 781 GAGTTTCGACGCGGAGCTTTCGAGAGATGACGAGCGGTACACAGAGAGAGCTGAG 840
DB 1496 GAGTTTCGACGCGGAGCTTTCGAGAGATGACGAGCGGTACACAGAGAGAGCTGAG 1555
QY 841 AACATGACCAAGCAGAGCTATCAAGAGATGACGAGTGAAGTGAAGTGAAGTGAAGTGAAG 900
DB 1556 AACATGACCAAGCAGAGCTATCAAGAGATGACGAGTGAAGTGAAGTGAAGTGAAGTGAAG 1615
QY 901 ATGAGAGACGAGAGACACCGGCTGCGGTGAGAGACAGCGGCTGGTGGAGACAGCGG 960
DB 1616 ATGAGAGACGAGAGACACCGGCTGCGGTGAGAGACAGCGGCTGGTGGAGACAGCGG 1675
QY 961 CGTGGCGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1020
DB 1676 CGTGGCGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1735
QY 1021 ACCGAGAACGAACTGACCGGCGAGCAGAGAGCGCGCTTTCCAAAGTTTGGAGACTAG 1080
DB 1736 ACCGAGAACGAACTGACCGGCGAGCAGAGAGCGCGCTTTCCAAAGTTTGGAGACTAG 1795
RESULT 5
ABL90067/c
ID ABL90067 standard; cDNA; 1083 BP.
XX
AC ABL90067;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 629.
XX
DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antituber;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200190304-A2.
XX
XX 29-NOV-2001.
XX

OY 181 CCGAGGGGGAAGGAGCCTGGAAATCCCAACCACTCCCTTGACAGCCAGGCGCTGCCA 240
DB 284 CCGGAGGGGGAAGGAGCCTGGAAATCCCAACCACTCCCTTGACAGCCAGGCGCTGCCA 343
OY 241 GAATCTAGCTGCTGAGAGAGGGCGAGAAAGGGCCGAATGGGAGCAGCTGTCGCTGCG 300
DB 344 GAATCTAGCTGCTGAGAGAGGGCGAGAAAGGGCCGAATGGGAGCAGCTGTCGCTGCG 403
OY 301 GGGCACTTCCCGCCCGCCGACAAAGTGGAAACCGAGCGCCGAGCGCCGAGCTGCTGCCCG 360
DB 404 GGGCACTTCCCGCCCGCCGACAAAGTGGAAACCGAGCGCCGAGCGCCGAGCTGCTGCCCG 463
OY 361 CTTGTGATGATCTCCGAGGCGCAGTAAGTTGGGGGCTCCTCGCCGAGGGGCGCAAG 417
DB 464 CTTGTGATGATCTCCGAGGCGCAGTAAGTTGGGGGCTCCTCGCCGAGGGGCGCAAG 520

RESULT 7
ABL63543
ID ABL63543 standard; DNA; 461 BP.

AC ABL63543;
XX
XX 15-MAY-2002 (first entry)
DT
XX
XX Breast cancer related gene sequence SFO ID NO:1880.
DE
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; kiltumour; cancerous;
KW cytoslastic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
OS
XX Homo sapiens.
XX
XX WO200194629-A2.
PN
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
XX 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.

PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI: 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 1880; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (1) of a signature gene set, where (1)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (1) has cytoslastic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
XX Sequence 461 BP; 105 A; 134 C; 138 G; 80 T; 4 other;

Query Match 32.9%; Score 355.2; DB 24; Length 461;
Best Local Similarity 95.7%; Pred. No. 3e-64; Mismatches 16; Indels 1; Gaps 1;
Matches 374; Conservative 0;

OY 1 ATGCCGAGCCATTCTTGTGCAAAATATCAACACCAAGCTCAACTAGCACTGTACAGGT 60
DB 72 ATGCCGAGCCATTCTTGTGCAAAATATCAACACCAAGCTCAACTAGCACTGTACAGGT 131
OY 61 GCTGCTGCTGTCCAGAGAGCTGAACCTTGAGCCGCCCGGAGGAGGCGGGT 120
DB 132 GCTGCTGCTGTCCAGAGAGCTGAACCTTGAGCCGCCCGGAGGAGGCGGGT 191
OY 121 CCCGAGAGAGACAGTAGTGTGCAATCGAGAGGTTCCCAATTGGTTCGCGGGG 180
DB 192 CCCGAGAGAGACAGTAGTGTGCAATCGAGAGGTTCCCAATTGGTTCGCGGGG 251
OY 181 CCGGAGGGGAGAGAGCCTGGAATCCCAACCACTCCCTTGCAACCCAGGCTGTGCA 240
DB 252 CCGGAGGGGAGAGAGCCTGGAATCCCAACCACTCCCTTGCAACCCAGGCTGTGCA 311
OY 241 GAATCTAGCTGCTGAGAGAGGGCGAGAAAGGGCCGAATGGGAGCAGCTGTCGCTGCG 300
DB 312 GAATCTAGCTGCTGAGAGAGGGCGAGAAAGGGCCGAATGGGAGCAGCTGTCGCTGCG 371
OY 301 GGGCACTTCCCGCCCGCCGACAAAGTGGAAACCGAGCGCCGAGCGAGCTGTCGCCAG 360
DB 372 -GGCACTTCCCGCCCGCCGACAAAGTGGAGGCCGAGCGCGAGCTGTCGCCAG 430
OY 361 CTTGTGATGATCTCCGAGGCGCAGTAAGTTGG 391

Db 431 CCTGTGATGACTCCGAGGCGAGTAGTGG 461
 |||
 RESULT 8
 AAV87316
 ID AAV87316 standard; cDNA: 342 BP.
 AC AAV87316;
 XX
 XX 27-APR-1999 (first entry)
 DT
 DE EST clone BP783.
 XX
 XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 KW
 XX Homo sapiens.
 OS
 PN M09845435-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US06954.
 XX
 PR 10-APR-1997; 97US-0835913.
 XX
 PA (GEMV) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX
 DR WPI: 1999-070076/06.
 XX
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 PS
 PS Claim 1; Page 526; 633pp; English.
 XX
 CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC therapy. The EST sequences are also stated to be useful for gene
 CC
 XX
 S0 Sequence 342 BP; 90 A; 79 C; 83 G; 89 T; 1 other;
 Query Match 12.8%; Score 138; DB 20; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1.9e-19;
 Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGCCGAGCCATTCTGTGATGATATCAACACAGCCTCAAACTAGCACTGTACAGT 60
 DB 196 ATGGCCGAGCCATTCTGTGATGATATCAACACAGCCTCAAACTAGCACTGTACAGT 255
 QY 61 GCTCTGCTGTGTCAGGAGAGCTGACCCCTGAGCCGCCGCGGAGAGCGGGTG 120
 DB 256 GCTCTGCTGTGTCAGGAGAGCTGACCCCTGAGCCGCCGCGGAGAGCGGGTG 315
 QY 121 CCCGAGGAGGACACTAGG 138
 |||

Db 316 CCCGAGGAGGACACTAGG 333
 |||
 RESULT 9
 AAH99667/C
 ID AAH99667 standard; cDNA: 997 BP.
 AC AAH99667;
 XX
 XX 16-OCT-2001 (first entry)
 DT
 DE Human protein encoding cDNA sequence SEQ ID NO:502.
 XX
 XX
 XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiac; central nervous system; vitruide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antineumatic; anaemia;
 KW antiaggregant; haemostatic; vulnery; antitumor; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder; ss.
 KW
 XX Homo sapiens.
 OS
 PN W0200153455-A2.
 XX
 PD 26-JUL-2001.
 XX
 PF 22-DEC-2000; 2000WO-US35017.
 XX
 PR 23-DEC-1999; 99US-0471275.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 XX
 PR 25-APR-2000; 2000US-0552317.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Drmanac RT;
 PI
 XX
 DR WPI: 2001-457603/49.
 XX
 DR P-PSDB: AAM25726.
 XX
 PT Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 PS
 PS Claim 1; Page 566-567; 1217pp; English.
 XX
 CC AAH99166 to AAH99904 encode the human proteins given in AAM2525 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; vitruide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antineumatic; antiaggregant; haemostatic; vulnery;
 CC antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and

CC neurological disorders.
XX
SQ Sequence 997 BP; 184 A; 323 C; 266 G; 224 T; 0 other;
Query Match 10.4%; Score 112.4; DB 22; Length 997;
Best Local Similarity 62.7%; Pred. No. 4.5e-14;
Matches 192; Conservative 0; Mismatches 11; Indels 3; Gaps 1;
QY 733 GGGGGTGGAGAGATGGGGCCAGCATGGATGGAGGGAGCCGACGATTTCGAC 792
DB 887 GGGGAAGTGTAGCCCGGACAGTGTATGGGGGGGCGACGCGCATGATTCACCCG 828
QY 793 CGGAGCTTCGAGACGTACGACGCTACACAGGAGCCTGCAGACATGAGCAG 852
DB 827 AAGGACTTCTCTAGACTTACGAACTTCCACACGAGAGCCTGCAGGGCCCGACAG 768
QY 853 CAGGAGCTCATCAAGAGTACCTGGAAGTGGAGAGTCCCTCCGCGCATGAGAGCAG 912
DB 767 CAGGAGCTGTGTGCGAGACTACCTGAGAGTGGAGAGGCGCTGTCCGAGCGAGAGAG 708
QY 913 AACACCGGCTGGCGCTGGAGAGCAACCGGCTGGCTGCCA---CGACGCGCTGTCCG 969
DB 707 ACTAGAGGCTGCGACAGCTGCAGGGCTGCACCGCCAGCATGCTCCGCGCAGGTGAG 648
QY 970 GAGCTGAGCTGGAGCTGAGACCGGCTGCGCGCAGAACCTCCAGCTGTGACCGAGAAC 1029
DB 647 GAGCTGCTGCTGCGAGAGTCCAGAGGCTCCGAGCCGAAACCGAGCGGCTTCCTCAGAGAAC 588
QY 1030 GAACGTG 1035
DB 587 CAGATG 582
RESULT 10
AAH12885/c
ID AAH12885 standard; cDNA; 567 BP.
XX
AC AAH12885;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (3'-primer) SEQ ID NO:9720.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3: SEQ ID 9720; 2537bp + CD ROM; English.
XX

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises at least 15 nucleotides, where the
CC oligonucleotide comprises the 5'-end sequence, where the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 567 BP; 182 A; 139 C; 87 G; 149 T; 10 other;
Query Match 7.7%; Score 82.8; DB 22; Length 567;
Best Local Similarity 76.7%; Pred. No. 5.2e-08;
Matches 122; Conservative 0; Mismatches 34; Indels 3; Gaps 2;
QY 922 CTGCGCTGGAGAGCAGCGCTGTGGTGGAGACCGCGGCTGGCGGAGTGGAGCTG 981
DB 539 CNGNTTCGTTGAGACANCGTTGGTGGGAAACCGGCTTTGGCGGAG-TGGAGTTG 481
QY 982 GAGCTGACCGCGCTGCGCGCGGAGACCTCCAGCTGCTGACCGAGAACGAACTGCACCG 1041
DB 480 GAGTTG--ACCGGTGGCGCGCGGAGAACTTCAGTTGTTGACGAGAAAGAAATTGCACCG 423
QY 1042 CAGCAGAGGAGCGCGCGCTTCCAGATTGGAGACTAG 1080
DB 422 CAGCAGAGGAGCGCGCGCTTCCAGATTGGAGACTAG 384
RESULT 11
AAV55831
ID AAV55831 standard; DNA; 799 BP.
XX
AC AAV55831;
XX
DT 18-NOV-1998 (first entry)
XX
DE Nucleotide sequence of the stabilising sequence-encoding insert.
XX
OS Fusion protein; stabilising polypeptide; proteolytic degradation;
XX resistance; half-life; autoimmune disease; inflammation; nitro drug;
XX IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
XX nitroreductase protein; enzyme therapy; produg therapy; protease;
XX cancer; pathological condition; ss.
XX
OS Epstein-barr virus.
XX
PN WO9822577-A1.
XX
PD 28-MAY-1998.
XX
PE 17-NOV-1997; 97WO-IB01508.
XX
PR 25-JUN-1997; 97US-0048945.
XX
PR 15-NOV-1996; 96US-0030986.
XX
PA (MASU/) MASUCCI M G.
XX
PI Masucci MG;
XX

Best Local Similarity 44.0%; Pred. No. 6e-05;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```

QY 406 GGGGGGGAAGAGAGTGGGACACGACACACACTGGGAAAGAAAAACATAAGAGA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 GGGCAGGAGCAGAGGGGCGAGAGCAGAGAGGGGCGAGGGGCGAGGGGCGAG 364
QY 466 CCGCCCTCCAAAGAAAGCGCATTTGAAACCTACTACAGCTGAAGTGGAAAGAA 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 GAGGGGCAGAGCAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGGGGCGAG 424
QY 526 AAAAAGTTGACGAGAAGACAGACCTTCGAGCTTCAAGATCCAGCCGAGATGTTCC 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 CAGACGACAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGGGGCGAGGGGCGAG 484
QY 586 AAGGGCCAGCCGCTCCGCTTATACACCAGCCAGTTCCTCATGATGATCAGACAG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 GGGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGGGGCGAGGGGCGAGGGGCGAG 544
QY 646 GAGGAGCCGATCTCAAAACCGGCTGTACTCCAGCCGGGCGCCGCAAAATCCAGAC 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 GAGGGGCAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 604
QY 706 ACCAGCGATGACGACTTCAATGGAAGAAGGGGCTGAGAGAGATGGCGCAGCGATG 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 CAGGACGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 664
QY 766 GGAAGGGGAGCGGACGCAATTTCTGCAAGCCGAGCTTCCGAGACGTAACGACGTT 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 GAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 724
QY 826 ACAGAGAGCCTGACAAATGAGCAGAGCAGGACCTCATCAAGAGTAACTGGAAGT 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 CAGGAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 784
QY 886 AAGTGCCTCTCGCGATGAGAGCAGAAACACCGGCTTGGAGAGCAACCGGCTG 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 GAGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAG 844
QY 946 GGTGGGAGAGAGCGCGCTGTGCGGAGCTGTGAGACTGGAGTCCGCGCTGCCCG 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 845 GAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 904
QY 1006 AACCTCCAGCTGTGACCGAGAACTGCAACCGCAGCAGAGAGAGAGAGCGCG 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 GGGCAGAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGGGGCGAGAGAG 958

RESULT 13
AAF82902
ID AAF82902 standard: DNA; 1926 BP.
XX
XX AAF82902:
XX
XX 29-JUN-2001 (first entry)
XX
XX EBV tethering protein EBNA1 encoding DNA.
XX
XX Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
XX Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
XX EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.
XX
XX Epstein-barr virus.
XX
XX
XX Key Location/Qualifiers
XX CDS 1..1926
XX /*tag= a
XX
XX MO200125484-A2.
XX
XX 12-APR-2001.
XX
XX 29-SEP-2000; 2000MO-US26908.

```

```

XX
PR 01-OCT-1999; 99US-0410399.
XX
XX (UNM1) UNIV MICHIGAN.
XX
XX Robertson ES; Cotter MA;
XX
XX WPI: 2001-281736/29.
XX
XX P-PSDB; AAB62332.
XX
XX A composition for use in gene therapy comprises an expression vector
XX PT that includes a nucleic acid sequence encoding a nucleic acid binding
XX protein
XX
XX Disclosure; Fig 9C; 60pp; English.
XX
XX The invention provides a composition comprising nucleic acid, histone H1
XX CC protein and expression vector operationally encoding a protein suitable
XX CC for tethering the nucleic acid to the histone H1 protein, where the
XX CC tethering protein is LANA. The composition is useful in aiding the
XX CC retention of the viral DNA in the host cell. The viral vector encodes a
XX CC protein suitable for tethering DNA to histone H1. Methods for screening
XX CC for compounds which are agonistic or antagonistic for the tethering of
XX CC viral proteins to histone H1 and DNA binding sites are useful for
XX CC developing the method of viral transfer. The composition has applications
XX CC to gene therapy, including the treatment of multiple sclerosis,
XX CC Parkinson's disease, Huntington disease and diabetes. The present
XX CC sequence represents the nucleotide sequence of the Epstein-barr virus
XX CC (EBV) tethering protein EBNA1.
XX
XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;
XX
XX

```

Query Match 6.3%; Score 68.4; DB 22; Length 1926;

Best Local Similarity 44.0%; Pred. No. 6e-05;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```

QY 406 GGGGGGGAAGAGAGTGGGACACGACAGACACTGGGAAAGAAAAACATAAGAGA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 GGGCAGGAGCAGAGAGGGGCGAGAGCAGAGAGGGGCGAGGGGCGAGGGGCGAG 364
QY 466 CCGCCCTCCAAAGAAAGCGCATTTGAAACCTACTACAGCTGAAGTGGAAAGAA 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 GAGGGGCAGAGCAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAG 424
QY 526 AAAAAGTTGACGAGAAGACAGACCTTCCAGCTTCAAGATCCGAGCCAGATGTTCC 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 CAGACGACAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 484
QY 586 AAGGGCCAGCCGCTCCGCTTATACACCAGCCAGTTCCTCATGATGATCAGACAG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 GGGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 544
QY 485 GAGGAGCCGAGATCTCAAAACCGGCTGTACTCCAGCGGCGCCGCAAAATCCAGAC 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 GAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 604
QY 706 ACAGAGGATGACGACTTCAATGGAAGAAGGGGCTGAGAGAGATGGGCGAGCGAT 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 CAGGAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 664
QY 766 GGAAGGGGAGCGGACGCAATTTCTGCAAGCCGAGCTTCCGAGACGTAACGAGCT 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 GAGGGGCAGAGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 724
QY 826 ACAGAGAGCCTGACAAATGAGCAGAGCAGGACCTCATCAAGAGTAACTGGAAGT 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 CAGGAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 784
QY 886 AAGTGCCTCTCGCGATGAGAGCAGAAACACCGGCTTGGAGAGCAACCGGCTG 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 GAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 844
QY 946 GGTGGGAGAGAGCGCGCTGTGCGGAGCTGTGAGACTGGAGTCCGCGCTGCCCG 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


CC protein-protein interactions between known proteins and for isolating,
CC cloning or characterizing unknown proteins. The method is also useful
CC for screening candidate bioactive agents that modulate the
CC protein-protein interaction between a bait protein and a test protein
CC and thus are useful for the identification of novel therapeutic drugs.
CC The present method employs mammalian cloning systems that are highly
CC stable, and designed to reduce the background signals frequently found
CC in other systems such as the yeast two-hybrid system. The current
CC sequence represents the coding sequence for the Epstein-Barr virus
CC nuclear antigen, the viral replication protein, which, on binding to the
CC Epstein-Barr origin of replication (see AI164274) on the bait vector,
CC allows vector replication to occur.
XX

SO Sequence 2580 BP; 632 A; 512 C; 1055 G; 381 T; 0 other;

Query Match 6.38; Score 68.4; DB 24; Length 2580;

Best Local Similarity 44.08; Pred. No. 6.3e-05;

Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

```
QY 406 GGGGCGAAGAGAGTGGGACAGACAGACGCTGGGGAAGAAAAACATAAGAGA 465
    ||| || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 688 GGGGAGAGAGCGAGAGAGGCGCAGAGACAGAGAGGGGCGAGAGGGGCGAG 747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 CGCCCGTCCAGAGAAAGCGGCATTGGAAACCGTACTACAAGCTGAACCTGGGAAGAGA 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 748 GAGGGGCGAGAGAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGGGG 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 526 AAAAAGTTCCAGACGAGAAACAGAGCCTTCGAGCTTCAANGATCCGAGCCGAGATTTGCC 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 808 CAGAGCAGAGAGAGAGGGGCGAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAG 867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 586 AAGGGCCAGCGGTCGGCCCTTATACACACGCAATTCATGATGATCAGCAGCAG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 868 GGGGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGGGGCGAG 927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 646 GAGGAGCGGATCTCAAAACCGGCTTACTTCCAGCGGGCCGCCCAATCCGACGAC 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 928 GAGGGCGAGAGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGGGG 987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 706 ACCAGCGATGACGACTTCATGGAAGAAGGGGCTAGAGAGATGGGGCAGCGATGGCATG 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 988 CAGGAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAG 1047
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 766 GAGGAGGAGCGACGAGATTCTTCGACGCGGACTTCTCGAGACTAGAGCGGTATCCAC 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1048 GAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGGGGCGAG 1107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 826 ACGGAGAGCTGCGAAGCATGACAGCAGAGACCTCATCAAGAGACTCTGGAACCTGAG 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1108 CAGGAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 1167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 886 AAGTGCCTCTCGCATGAGAGAGAGAACACCGGCTGCGGCTGAGAGCAGACGGGCTG 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1168 GAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 1227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 946 GGTGCGCAGCAGCGGCTGTGCGGGAGCTGAGCTGAGCTGACCGGCTGCGCGCGAG 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 GAGGGGCGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 1287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1006 AACCTCAGCTGCTGACCGAGAACGAACTCCACCAGCAGAGAGCGGCCG 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 GGGCAGAGAGCAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCAG 1341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: June 18, 2003, 01:10:55
Job time : 321 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 21:37:42 ; Search time 1820 Seconds
(without alignments)
9610.509 Million cell updates/sec

Title: US-09-972-758a-1
Perfect score: 1080
Sequence: 1 atgagcagacattctgtc.....tttcaagtttgagactg 1080

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	68.1	876	9	AL552998
2	727.8	67.4	903	14	BO221641
3	707.2	65.5	982	12	BF984049
4	694.6	64.3	771	13	BI871190
5	652.8	60.4	852	13	BI091005
6	639.2	59.2	797	12	BG819109

Result No.	Score	Query Match	Length	DB ID	Description
7	628	58.1	1001	9	AL578242
8	615.2	57.0	871	12	BG291150
9	598	55.4	750	10	AV726891
10	585	54.2	949	12	BG165450
11	580.6	53.8	1294	13	BM474968
12	579	53.6	986	9	AL576112
13	572.6	53.0	1080	13	BM474949
14	552.2	51.1	557	13	BM129736
15	539.2	49.9	745	10	BE547281
16	525.2	48.6	933	12	BF674587
17	509.8	47.2	843	12	BF026064
18	498.2	46.1	907	12	BF181535
19	494.2	45.8	810	12	BF607249
20	488.4	45.2	563	10	AM665096
21	487.2	45.1	994	9	AL555845
22	486.4	45.0	762	12	BF167480
23	483.8	44.8	999	13	BI411874
24	477.4	44.2	482	9	AI948609
25	476.6	44.1	566	10	AM962738
26	471.2	43.6	547	9	AI796944
27	466.4	43.2	666	12	BG644490
28	438.4	40.6	907	13	BI412550
29	436.4	40.4	442	9	AI948617
30	432	40.0	638	9	AA546538
31	402	37.2	442	12	BF082480
32	391.6	36.3	462	10	AV728223
33	375.4	34.8	485	9	AI811680
34	372.2	34.5	450	10	AM166154
35	370.2	34.3	459	12	BF446806
36	368.6	34.1	577	13	BM440009
37	365.6	33.9	819	14	BO222807
38	363.6	33.7	418	14	W34238
39	360.6	33.4	540	10	AM143910
40	358	32.9	366	10	AV708410
41	355.2	32.1	461	14	N21081
42	342.6	31.7	980	13	BI526654
43	331	30.6	734	12	BG122671
44	330	30.6	400	9	AL362139
45	330	30.6	480	9	AA928789

ALIGNMENTS

RESULT 1
LOCUS AL552998 876 bp mRNA
DEFINITION AL552998 LTI_NFL006.PL2 Homo sapiens CDNA clone CS0D1072YL06 5
prime, mRNA sequence.
ACCESSION AL552998
VERSION AL552998.1 GI:12892417
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 876)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

1. 876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_NFL006.PL2"
/issue_type="Placenta"
/note="Vector: PCWVS-PORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@life.com
 http://fulllength.invitrogen.com"

BASE COUNT 221 a 235 c 266 g 152 t 2 others

Query Match 68.1%; Score 736; DB 9; Length 876;
 Best Local Similarity 99.3%; Pred. No. 4.4e-140;
 Matches 747; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

1 ATGGCGAGCATTCTTGTGCAATATATACACAGCCTCAACATAGCACTGTACAGT 60
 126 ATGGCGAGCATTCTTGTGCAATATATACACAGCCTCAACATAGCACTGTACAGT 185
 61 GCTGCTGCTGTCAGAGAGAGTGAACCTGAGCGCCCGGCGGAGAGCGGGTG 120
 186 GCTGCTGCTGTCAGAGAGAGTGAACCTGAGCGCCCGGCGGAGAGCGGGTG 245
 121 CCGGAGAGAGAGAGAGTGAACCTGAGCGCCCGGCGGAGAGCGGGTG 180
 246 CCGGAGAGAGAGAGTGAACCTGAGCGCCCGGCGGAGAGCGGGTG 305
 181 CCGGAGAGAGAGAGTGAACCTGAGCGCCCGGCGGAGAGCGGGTG 240
 306 CCGGAGAGAGAGAGTGAACCTGAGCGCCCGGCGGAGAGCGGGTG 365
 241 GATCTAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 366 GATCTAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425
 301 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 426 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
 361 CTTTGTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 486 CTTTGTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
 421 TGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 545 TGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
 481 AAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 605 AAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 664
 541 AAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 665 AAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 724
 601 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 725 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 784
 661 AAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 785 AAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 844
 721 TTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 752
 845 TTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876

RESULT 2
 BQ221641 903 bp mRNA linear EST 02-MAY-2002
 LOCUS AGENCOURT_7552882 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6058110
 DEFINITION 5', mRNA sequence.

ACCESSION BQ221641 GI:20403041
 VERSION BQ221641.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mhc.nhl.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: DCTD/DRP/Gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM1332 row: k column: 07
 High quality sequence stop: 644.
 Location/Qualifiers

FEATURES
 source 1..903
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6058110"
 /clone_lib="NIH-MGC_68"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lung; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

BASE COUNT 217 a 259 c 309 g 115 t 3 others

Query Match 67.4%; Score 727.8; DB 14; Length 903;
 Best Local Similarity 97.9%; Pred. No. 2.1e-138;
 Matches 759; Conservative 0; Mismatches 12; Indels 4; Gaps 2;

151 GCGTCCCGCCAGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 210
 1 GCGTCCCGCCAGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
 211 CCACCTCCCTTGACAGACCCAGGCTGTGTCAGATCTAGTGTGCTGAGAGAGGCGGAGAG 270
 61 CCACCTCCCTTGACAGACCCAGGCTGTGTCAGATCTAGTGTGCTGAGAGAGGCGGAGAG 120
 271 GCGCAGAAATGGGAGAGCTGTCGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 330
 121 GCGCAGAAATGGGAGAGCTGTCGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 331 CCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 390
 181 CCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 241 GGGGCTCTGCGCGGAGGGGGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 391 GGGGCTCTGCGCGGAGGGGGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 450
 241 GGGGCTCTGCGCGGAGGGGGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 451 AAAAAACATAGAGAGCGCCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 510
 301 AAAAAACATAGAGAGCGCCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 511 AACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570
 361 AACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 571 GCGGAGATGTGCGCAAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 630
 421 GCGGAGATGTGCGCAAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

QY 631 GATGATCACGACGAGAGAGCCGGATCTCAAAACCGGCTGTACTCCAGGGGGCGCC 690
 Db 481 GATGATCACGACGAGAGAGCCGGATCTCAAAACCGGCTGTACTCCAGGGGGCGCC 540
 QY 691 GCCAATCCGACGACACGAGATGACGACTTTCATGAGAAAGGGGGTGGAGAGATGG 750
 Db 541 GCCAATCCGACGACACGAGATGACGACTTTCATGAGAAAGGGGGTGGAGAGATGG 600
 QY 751 GGCAGCAGTGGGATGGAGGAGCGGCGAGGATTTTCGACGGGAGCTTTCGGAGAGC 810
 Db 601 GGCAGCAGTGGGATGGAGGAGCGGCGAGGATTTTCGACGGGAGCTTTCGGAGAGC 660
 QY 811 TACGAGCGGTACACGACGCG -AGAGCTTCAGACATGAGCAAGCAGGAGCTCATCAAGA 869
 Db 661 TACGAGCGGTACACGACGCGAGAGCTTCAGACATGAGCAAGCAGGAGCTCATCAAGA 720
 QY 870 GTACCTGGAACTGGAGAAGTGC---CTCTCGCGCGATGGAGAGCAGCAACCGG 921
 Db 721 GTACCTGGAACTGGAGAAGTGCCTCTCTCGCGCGATGGAGAGCAGCAACCGG 775
 RESULT 3
 BP984049 982 bp mRNA linear EST 23-JAN-2001
 LOCUS 602306989F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4398238 5',
 DEFINITION mRNA sequence.
 ACCESSION BP984049
 VERSION BP984049.1 GI:12386861
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH_MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10100 row: a column: 23
 High quality sequence stop: 691.
 FEATURES
 Source Location/Qualifiers
 1..982
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4398238"
 /clone_lib="NIH_MGC_88"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: pCMV-SPORT6;
 Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
 oligo-dt primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."
 BASE COUNT 261 a 235 c 320 g 166 t
 ORIGIN
 Query Match 65.5%; Score 707.2; DB 12; Length 982;
 Best Local Similarity 99.4%; Pred. No. 3.3e-134;
 Matches 720; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
 QY 357 CCACGCTTGATGATCGGACGCGCACTAGTTGGGGCTCTGCCGACGAGGGCGCAAGA 416
 Db 7 CCACGCTTGATGATCGGACGCGCACTAGTTGGGGCTCTGCCGACGAGGGCGCAAGA 66
 QY 417 GGAGTGGGAGACGACGACGACGCTGGGGAAGAAAAACATAGAGAGAGCCGTCCAA 476

Db 67 GGAGTGGGAGACGACGACGACGCTGGGGAAGAAAAACATAGAGAGCGCCGTCCAA 126
 QY 477 GAAGAGCGGCTTGGAAACCGTACTACAAAGCTGAACCTGGGAAGAGAAAAAGTTGCA 536
 Db 127 GAAGAAACCGGCAATGGGAACCGTACTACAAAGCTGAACCTGGGAAGAGAAAAAGTTGCA 186
 QY 537 CGAAGAAACAGAGGCTTCGAGCTTCAGAGGATCCGAGGCGAGATGTTGCCAAGGCCAGCC 596
 Db 187 CGAAGAAACAGAGGCTTCGAGCTTCAGAGGATCCGAGGCGAGATGTTGCCAAGGCCAGCC 246
 QY 597 GGTGCGGCCCTATTAACCAACGACGATTCCTCATGTGATGATCAAGCAACAGAGAGAGCCGA 656
 Db 247 GGTGCGGCCCTATTAACCAACGACGATTCCTCATGTGATGATCAAGCAACAGAGAGAGCCGA 306
 QY 657 TTCGAAGAACCGGCTGTACTCCAAAGCGGGGCGCCGCAATTCGACAGACACCGAGATGA 716
 Db 307 TTCGAAGAACCGGCTGTACTCCAAAGCGGGGCGCCGCAATTCGACAGACACCGAGATGA 366
 QY 717 CGACTTCATGAGAAAGAGGGGCTGAGAGGATGGGGCAGCGATGGAGGGGAGCG 776
 Db 367 CGACTTCATGAGAAAGAGGGGCTGAGAGGATGGGGCAGCGATGGAGGGGAGCG 426
 QY 777 CAGCAGTTTCTGACAGCGGAGCTTCTCGAGACGTAGCAGCGGTAACCAACAGAGAGCCT 836
 Db 427 CAGCAGTTTCTGACAGCGGAGCTTCTCGAGACGTAGCAGCGGTAACCAACAGAGAGCCT 486
 QY 837 GCAGAACATGAGCAAGCAGAGGCTGATCAAGAGCTGAACTGAGAGAGTCCCTCTC 896
 Db 487 GCAGAACATGAGCAAGCAGAGGCTGATCAAGAGCTGAACTGAGAGAGTCCCTCTC 546
 QY 897 GCGCATGAGAGACGAGCAACCAACCGCTCCGCTGAGAGCAAGCGGTGGGTGGCGAGCA 956
 Db 547 GCGCATGAGAGACGAGCAACCAACCGCTCCGCTGAGAGCAAGCGGTGGGTGGCGAGCA 606
 QY 957 CGCGGTGTCGGGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCT 1016
 Db 607 CGCGGTGTCGGGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCT 666
 QY 1017 GCTACCCGAGAACGAACTGCAACCGCAGCAGAGAGCGCCGCTTCCAAAGTTGGAGA 1076
 Db 667 GCTACCCGAGAACGAACTGCAACCGCAGCAGAGAGCGCCGCTTCCAAAGTTGGAGA 725
 QY 1077 CTAG 1080
 Db 726 CTAG 729
 RESULT 4
 B1871190 771 bp mRNA linear EST 11-OCT-2001
 LOCUS 603395081F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5404724 5',
 DEFINITION mRNA sequence.
 ACCESSION B1871190
 VERSION B1871190.1 GI:16044865
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH_MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Db 365 GGGGCGCGATGATGGAGGGAGCGCAGAG-TTCTGACGGGGAGCTTCTGAGGA 423
 Oy 809 CGTACAGACGGTACACACAGAGAGCTCTGAGACATGAGACAGAGAGCTCATCAAG 868
 Db 424 CGTACAGAGGGTACACACAGAGAGCTCTGAGACATGAGACAGAGAGCTCATCAAG 483
 Oy 869 AGTACCTGGAAGTGGAGAGTGGTGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 928
 Db 484 AGTACCTGGAAGTGGAGAGTGGTGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 543
 Oy 929 TGGAGAGCAAGCGCTGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 988
 Db 544 TGGAGAGCAAGCGCTGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 603
 Oy 989 ACCGGCTGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1048
 Db 604 ACCGGCTGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
 Oy 1049 AGCGAGCGCGCTTCCAGTTTGGAGACTAG 1080
 Db 663 AGCGAGCGCGCTTCCAGTTTGGAGACTAG 694

RESULT 6
 Bg819109 797 bp mRNA linear EST 22-MAY-2001
 LOCUS Bg819109 602781202P1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4932025
 DEFINITION 5', mRNA sequence.

ACCESSION Bg819109
 VERSION Bg819109.1 GI:14166696
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 797)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE Unpublished (1999)
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rstraus@nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: LLAM10857 row: c column: 02
 High quality sequence stop: 797.

FEATURES

source
 1..797
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4932025"
 /clone_lib="NCI CGAP_Brn67"
 /tissue_type="anaplastic Oligodendroglioma with 1p/19q loss"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Brain; Vector: PCMV-SPOK6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
 BASE COUNT 207 a 223 c 248 g 119 t
 ORIGIN

Query Match 59.2%; Score 639.2; DB 12; Length 797;
 Best Local Similarity 96.5%; Pred. No. 2.4e-120;
 Matches 708; Conservative 0; Mismatches 18; Indels 8; Gaps 5;
 Oy 1 ATGCCGAGCCATTCTTGTGATATCAACACAGCGCTCAACTGACACTGACGAT 60
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 69 ATGCCGAGCCATTCTTGTGATATCAACACAGCGCTCAACTGACACTGACGAT 128
 Oy 61 GCTGCTGCTTCCAGAGAGAGCTGAACCTGAGCGCCGCCAGCGCGAGAGAGAGAGAG 120
 Db 129 GCTGCTGCTTCCAGAGAGAGCTGAACCTGAGCGCCGCCAGCGCGAGAGAGAGAGAG 188
 Oy 121 CCCAGAGAGAGAGTGTAGTGGCAATGAGAGAGCTTCCCGAGTGGGTGGCCGTCCGGG 180
 Db 189 CCCAGAGAGAGAGTGTAGTGGCAATGAGAGAGCTTCCCGAGTGGGTGGCCGTCCGGG 248
 Oy 181 CCCAGAGAGAGAGAGCTTGAATCCCAACACCTTCCCTTGGAGAGAGAGAGAGAGAG 240
 Db 249 CCCAGAGAGAGAGAGCTTGAATCCCAACACCTTCCCTTGGAGAGAGAGAGAGAGAG 308
 Oy 241 GAATCTAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 Db 309 GAATCTAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 368
 Oy 301 GAGGACTTCCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 Db 369 GAGGACTTCCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
 Oy 361 CCTGTGATGACTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 Db 429 CCTGTGATGACTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
 Oy 421 TGGGGACAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 Db 489 TGGGGACAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 548
 Oy 481 AAGCGGCTTGGAAACCGTACTACAACTGAACTGGGAAGAGAGAGAGAGAGAGAGAGAG 538
 Db 549 AAGCGGCTTGGAAACCGTACTACAACTGAACTGGGAAGAGAGAGAGAGAGAGAGAGAG 607
 Oy 539 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
 Db 608 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667
 Oy 599 TCGCGCGCTATTAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
 Db 668 TCGCGCGCTATTAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 727
 Oy 658 CTCAAAAACCGCGCTGAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
 Db 728 CTCAAAAACCGCGCTGAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 783
 Oy 718 GACTTCATGAGAGA 731
 Db 784 GACTTCATGAGAGA 797

RESULT 7
 AL578242/c 1001 bp mRNA linear EST 16-FEB-2001
 LOCUS AL578242 LTI.NFL006.PL2 Homo sapiens cDNA clone CSDDK02YF07 3
 DEFINITION prime, mRNA sequence.

ACCESSION AL578242
 VERSION AL578242.1 GI:12942132
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 1001)
 AUTHORS L.L.W.B., Gruber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1..1001
 Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK002YR07"
/clone_1ib="LTL_NFL006_P12"
/tissue_type="Placenta"
/Note="Vector: PCWSPORT 6; Site:1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCWSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact: Feng Liang Life Technologies,
Rockville, Maryland 20850, USA Fax: (1) 301 610 8371
Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com"
BASE COUNT      218 a      289 c      271 t      5 others
ORIGIN

```

```

Query Match      58.1%; Score 628; DB 9; Length 1001;
Best Local Similarity 98.8%; Pred. No. 4.6e-118;
Matches 639; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

434 AGAGACAGCTGGGAGAGAAAAACATAGAGAGCCGCTCCAAAGAGAGCGGATTGGA 493
1001 AGAGACAGCTGGGAGAGAAAAACATAGAGAGAGCCGCTCCAAAGAGAGCGGATTGGA 942
494 AACCTACTACAGCTGAGCTGGAGAGAGAAAAAGTTGAGAGAGAAACAGAGCTTC 553
941 AACCTACTACAGCTGAGCTGGAGAGAGAAAAAGTTGAGAGAGAAACAGAGCTTC 882
554 GAGCTTCAAGAGTCCGAGCCGAGATGTTGGCAAGGGCCGAGCGGTGCGCCCTATACA 613
881 GAGCTTCAAGATTCGAGCCGAGATGTTGGCAAGGGCCGAGCGGTGCGCCCTATACA 822
614 CCAGCAGTTCCTATGATGATCAGACAGAGAGAGCGGATCTCAAAACCGGCTG 673
821 CCAGCAGTTCCTATGATGATCAGACAGAGAGAGCGGATCTCAAAACCGGCTG 762
674 ACTTCAAGCGGGCGCGCCCAAAATCCGAGACACACGAGCTGACGATTCATGGAAG 733
761 ACTTCAAGCGGGCGCGCCCAAAATCCGAGACACACGAGCTGACGATTCATGGAAG 702
734 GGGGTGAGAGATGGGGGAGCGGATGGATGGAGAGGAGCGACGAGTTTCGACG 793
701 GGGGTGAGAGATGGGGGAGCGGATGGATGGAGAGGAGCGACGAGTTTCGACG 642
794 GGGACTTCTCGAGAGATGACGAGCTGACGAGAGAGCTGCGAAGCATGAGCAAGC 853
641 GGGACTTCTCGAGAGATGACGAGCTGACGAGAGAGAGCTGCGAAGCATGAGCAAGC 582
854 AGGAGCTCATCAAGAGATGAGTGAAGCTGAGAGAGAGCTGCGGAGATGGAGAGGA 913
581 AGGAGCTCATCAAGAGATGAGTGAAGCTGAGAGAGAGCTGCGGAGATGGAGAGGA 522
914 ACAACGGGCTGGGGCTGGAGAGAGAGAGCTGGGTGGCGAGAGCGGCTGTGGGGAGC 973
521 ACAACGGGCTGGGGCTGGAGAGAGAGAGCTGGGTGGCGAGAGCGGCTGTGGGGAGC 462
974 TGGAGCTGAGAGTGAACCGGCTGGCGGCGGAGAGAGCTCAGTGTGTGACCGAGAGC 1033
461 TGGAGCTGAGAGTGAACCGGCTGGCGGCGGAGAGAGCTCAGTGTGTGACCGAGAGC 403
403 TGGAGCTGAGAGTGAACCGGCTGGCGGCGGAGAGAGCTCAGTGTGTGACCGAGAGC 1080
402 TGCACCGGAGAGAGAGAGAGCGGCTTTCACAACTTTGAGAGACTAG 356

```

```

RESULT 8
LOCUS      BG291150
DEFINITION 6038718271 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4515812 5',
            mRNA sequence.
ACCESSION  BG291150

```

```

VERSION      BG291150.1 GI:13048811
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cga@ds.fremail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLNL0406 row: d column: 21
              High quality sequence stop: 730.

```

```

FEATURES
source
Location/Qualifiers
1..871

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4515812"
/clone_1ib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: bladder; Vector: PCW-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT      210 a      230 c      324 g      107 t
ORIGIN

```

```

Query Match      57.0%; Score 615.2; DB 12; Length 871;
Best Local Similarity 91.9%; Pred. No. 1.9e-115;
Matches 750; Conservative 0; Mismatches 53; Indels 13; Gaps 9;

176 GGGGCGGAGAGGGGAGAGAGCTGGAATCCAAACCTCCTTCGACAGCCAGGCTT 235
9   CCGGGGGGAGAGGGGAGAGAGCTGGAATCCAAACCTCCTTCGACAGCCAGGCTT 68
236 GTCCAGATATGAGTGTGAGAGAGAGGAGAGAGGAGAGAGAGTGGAGAGTCTGCG 295
69 GTCCAGATATGAGTGTGAGAGAGAGGAGAGAGGAGAGAGAGTGGAGAGTCTGCG 128
296 CTGGCGGAGAGTTCGCGCCCGGAGAGAGTGGAGAGAGAGAGAGAGAGTGGAGAG 355
129 CTGGCGGAGAGTTCGCGCCCGGAGAGAGTGGAGAGAGAGAGAGAGAGTGGAGAG 188
356 CCCAGCTTCTCATGAGTCTCGAGAGAGAGTGGAGAGAGAGTGGAGAGAGTGGAGAG 415
189 CCCAGCTTCTCATGAGTCTCGAGAGAGAGTGGAGAGAGAGTGGAGAGAGTGGAGAG 247
416 AGGAGTGGAGAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAGAGTGGAGAGAGAG 475
248 AGGAGTGGAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAGAGTGGAGAGAGAGAG 307
476 AGAAGAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAG 535
308 AGAAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAGAG 366
536 ACAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAGAGAG 595
367 ACAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAGAGTGGAGAGAGAGAGAGAGAGAG 426
596 CGGTGCGGCTTATTAACACAGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 655
427 CGGTGCGGCTTATTAACACAGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 486

```

Query Match	55.48;	Score 598;	DB 10;	Length 750;
Best Local Similarity	88.0%;	Pred. No. 5.8e-112;		
Matches 680; Conservative	0;	Mismatches 39;	Indels 54;	Gaps 1
QY	300	CGGGGACCTTCCCGCGCGCGGAGTGTGAGAACGACGCCCGGAGGCGCAGATGTCGTGCGCCA		355
Db	1	CGGGGACTTCCCGCGCGCGGAGTGTGAGAACGACGCCCGGAGGCGCAGATGTCGTGCGCCA		60

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 949)	NIH-MGC http://mgc.nci.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
				Contact: Robert Strausberg, Ph.D. Email: cgapds-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:

http://image.llnl.gov
Plate: LLAM0250 row: k column: 07
High quality sequence stop: 622.
Location/Qualifiers

FEATURES

source

1. 949
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4456062"
/clone_lib="NIH_MGC_89"
/tissue_type="hypertrophied, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

BASE COUNT 243 a 257 c 335 g 114 t
ORIGIN

Query Match 54.2%; Score 585; DB 12; Length 949;
Best Local Similarity 90.8%; Pred. No. 2.7e-109;
Matches 747; Conservative 0; Mismatches 60; Indels 16; Gaps 11;

228 CCAGGCGCTGTCAGATCTAGCTGCTAGAGAGAGGCGCAAGAGGCGCAAGTGGGAGCA 287
2 CCAGGCGCTGTCAGATCTAGCTGCTAGAGAGAGGCGCAAGAGGCGCAAGTGGGAGCA 61
288 CTCGTCGCTGCGGCGGCACTTCCGCGCGCGCAAGAGTGGACCGCGCGCGCGCA 347
62 CTCGTCGCTGCGGCGGCGGCACTTCCGCGCGCGCGCAAGAGTGGACCGCGCGCGCA 120
348 GCTGTCGCGGCGGCGGCGGCACTTCCGCGCGCGCGCAAGAGTGGACCGCGCGCGCA 407
121 GCTGTCGCGGCGGCGGCGGCACTTCCGCGCGCGCGCAAGAGTGGACCGCGCGCGCA 179
408 GGGCGAAGAGAGTGGGAGACAGCAGCAGAGACAGTGGGAGAGAGAGAGAGAGAGC 467
180 GGGCGAAGAGAGTGGGAGACAGCAGCAGAGACAGTGGGAGAGAGAGAGAGAGAGC 239
468 CCCGTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 527
240 CCCGTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 238
528 AAGGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 587
299 AAGGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 358
588 GGGCGAAGAGAGTGGGAGACAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 647
359 GGGCGAAGAGAGTGGGAGACAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 418
648 GAGAGCGGATCTCAAAACCGGCTGATCTCAAGCGGCGCGCGCAAGTGGGAGAGC 707
419 GAGAGCGGATCTCAAAACCGGCTGATCTCAAGCGGCGCGCGCAAGTGGGAGAGC 476
708 CAGGATGACGACTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 766
477 CAGGATGACGACTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 536
767 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 824
537 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 595
825 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 880
596 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 655
881 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 938
656 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 715
939 GGGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 997

Db 716 GCGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 775
Qy 998 GCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1040
Db 776 GCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 818

RESULT 11
BM474968
LOCUS
DEFINITION
AGNCOURT_6476735 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562828
5', mRNA sequence.
BM474968
VERSION
BM474968.1 GI:18524010
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 1294)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@strausberg.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12292 row: n column: 13
High quality sequence stop: 507.
Location/Qualifiers

FEATURES

source

1. 1294
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5562828"
/clone_lib="NIH_MGC_88"
/tissue_type="Duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-Sport6;
Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC library."

BASE COUNT 263 a 380 c 407 g 241 t 3 others
ORIGIN

Query Match 53.8%; Score 580.6; DB 13; Length 1294;
Best Local Similarity 89.8%; Pred. No. 2.2e-108;
Matches 681; Conservative 0; Mismatches 65; Indels 12; Gaps 5;

Qy 89 CTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 148
Db 1 CTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
Qy 149 GAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 208
Db 61 GAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Qy 209 AACCACTCCCTTGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 268
Db 121 AACCACTCCCTTGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180
Qy 269 AGGCGCAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 328
Db 181 AGGCGCAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Qy 329 AACCGAGCGGCGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 388
Db 241 AACCGAGCGGCGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 300

QY 389 TGGGGGCTCTGCGCAGAGGGGCGAGAGAGTGGGGACACAGCAGACAGTGGGGA 448
 Db 301 TGGGGGCTCTGCGCAGAGGGGCGAGAGAGTGGGGACACAGCAGACAGTGGGGA 360
 QY 449 AGAAAAACATAGAGAGCGCCGTCACAAAGAAAGCGGATTTGAAACCGTACTACAAGC 508
 Db 361 AGAAAAACATAGAGAGCGCCGTCACAAAGAAAGCGGATTTGAAACCGTACTACAAGC 420
 QY 509 TGAAGTGGAGAGAGAGAAAAAGTTGACGAGAAACAGAGCTTCGAGCTTCAAGGATCC 568
 Db 421 TGACCTGGAGAGAGAGAAAAAGTTGACGAGAAACAGAGCTTCGAGCTTCAAGGATCC 480
 QY 569 GAGCGAGATGTTGCGCAGAGGGCGAGCGGTCGCGCCCTATACACACAGCAG--TTCCCTC 627
 Db 481 GAGCGAGATGTTGCGCAGAGGGCGAGCGGTCGCGCCCTATACACACAGCAGTTTCTC 540
 QY 628 ATGATGATCAGCAGCAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGCGGG-- 685
 Db 541 ATGATGATCAGCAGCAGAGAGAGCGGATCTCCNAACCGGCTGTACTCCAGCGGGC 600
 QY 686 CCGCGCGCAATTC--GACGACACGAGC--GATGACGACTTCATG-----AAGAGGGG 736
 Db 601 CCGCGCGCAATTCGCGGACGACACGCGCATTTGACCATTCATGGAGAACAGGGGGGT 660
 QY 737 GTGAGAGAGATGGGGGCGAGCGATGGATGGAGGGGAGCGGACGAGTTCTCGACGCGG 796
 Db 661 GGAAGAGAGATGGGGGCGAGCGCATGGGGAATGGGGAAGGGGAGACGCGCGCCCAATT 720
 QY 797 ACTTCTGGAGACGTACGAGCGGTACACACGAGGAGC 834
 Db 721 TTTTCTTGACGCGGGGGAACCTTTCTTGGAAGAAC 758
 RESULT 12
 AL576112/ 986 bp mRNA linear EST 16-FEB-2001
 LOCUS AL576112 LTI_NFL006_PL12 Homo sapiens cDNA clone CS01072YL06 3
 DEFINITION AL576112 prime, mRNA sequence.
 ACCESSION AL576112
 VERSION AL576112.1 GI:12937932
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 986)
 AUTHORS Li, W. B., Gruber, C., Jesse, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 Location/Qualifiers
 1..986
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS01072YL06"
 /clone_1kb="LTI_NFL006_PL12"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
 BASE COUNT 216 a 284 c 217 g 265 t 4 others

Query Match 53.6%; Score 579; DB 9; Length 986;
 Best Local Similarity 96.7%; Pred. No. 4,5e-108;
 Matches 610; Conservative 2; Mismatches 17; Indels 2; Gaps 2;
 QY 450 GAAAAACATAGAGAGCGCCGTCACAAAGAAAGCGGATTTGAAACCGTACTACAAGC 509
 Db 985 GGAAGAAAAACATAGAAACCGCCGTCACAAAGAAAGCGGATTTGAAACCGTACTACAAGC 927
 QY 510 GAAGTGGAGAGAGAGAAAAAGTTGACGAGAAACAGAGCTTCGAGCTTCAAGGATCC 569
 Db 926 GACCTGGAGAGAGAGAAAAAGTTGACGAGAAACAGAGCTTCGAGCTTCAAGGATCC 867
 QY 570 AGCGAGATGTTGCGCAGAGGGCGAGCGGTCGCGCCCTATACACACAGCAGTTCTCAT 629
 Db 866 AGCGAGATGTTGCGCAGAGGGCGAGCGGTCGCGCCCTATACACACAGCAGTTCTCAT 807
 QY 630 GGATGATCAGACAGCAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAAGCGGCGC 689
 Db 806 GGATGATCAGACAGCAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAAGCGGCGC 747
 QY 690 CCGCAATTCGACGACACACAGCATGACGACTTCATGAAAGAGGGGTGACGAGATGG 749
 Db 746 CCGCAATTCGACGACACACAGCATGACGACTTCATGAAAGAGGGGTGACGAGATGG 687
 QY 750 GCGCAGCGATGGATGGGAGGGGAGCGGACGAGGATTTCTGACGCGGATCTTCGAGAC 809
 Db 686 GCGCAGCGATGGATGGGAGGGGAGCGGACGAGGATTTCTGACGCGGATCTTCGAGAC 627
 QY 810 GTACGAGCGGTACCCACAGAGAGCGCTTCAGAAACATGAGCAGAGAGCTCATCAAGA 869
 Db 626 GTACGAGCGGTACCCACAGAGAGCGCTTCAGAAACATGAGCAGAGAGCTCATCAAGA 567
 QY 870 GTACTGGAAGTGGAGAGTGCCTCTCGCGCATGAGAGAGAGAAACCGGCTGCGGCT 929
 Db 566 GTACTGGAAGTGGAGAGTGCCTCTCGCGCATGAGAGAGAGAAACCGGCTGCGGCT 507
 QY 930 GGAAGCAGACCGGCTGGGTCGCGACGCGCGGTGCGGAGCTGAGCTGAGCTGGA 989
 Db 506 GGAAGCAGACCGGCTGGGTCGCGACGCGCGGTGCGGAGCTGAGCTGAGCTGGA 447
 QY 990 CCGGCTGCGCGCGCAGAGACTCCAGCTGCTGACCGAGAGAACTGACCGAGCAGGA 1049
 Db 446 CC-GCTGCGCGCGAGAGACTCCAGCTGCTGACCGAGAGAACTGACCGAGCAGGA 388
 QY 1050 GCGAGCGCGCGCTTCCAAAGTTTGAGAGACTAG 1080
 Db 387 GCGAGCGCGCGCTTCCAAAGTTTGAGAGACTAG 357
 RESULT 13
 BM474949 1080 bp mRNA linear EST 05-FEB-2002
 LOCUS BM474949
 DEFINITION AGENCOURT_6476734 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562756
 ACCESSION BM474949
 VERSION BM474949.1 GI:18523991
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1080)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-femail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be

Db 61 AAAATTGACAGAAACAGAGCTTCGAGCTTCAAGATCCGACCCGAGATGTCGCA 120
OY 587 AGGCGCAGCCGCTCCGCCCTTAACACACGACGATTCCTCATATGATATCAGACGAG 646
Db 121 AGGGCCAGCCGCTCCGCCCTTAACACACGACGATTCCTCATATGATATCAGACGAG 180
OY 647 AGGACCCGATCTCAAAACCGGCCCTGTACTCCAGCGGCGCCGCAAAATCCGACGACA 706
Db 181 AGGACCCGACCTCAAAACCGGCCCTGTACTCCAGCGGCGCCGCAAAATCCGACGACA 240
OY 707 CCAGGATGACGACTTCAATGGAAGAGGGGCTGAGAGAGATGGGGCAACCGATGGATGG 766
Db 241 CCAGGATGACGACTTCAATGGAAGAGGGGCTGAGAGAGATGGGGCAACCGATGGATGG 300
OY 767 GAGGGAGCGCAGCAGATTTCTGACAGCGGACTTCCGAGACGTACAGAGCGGTACCCACA 826
Db 301 GAGGGAGCGCAGCAGATTTCTGACAGCGGACTTCCGAGACGTACAGAGCGGTACCCACA 360
OY 827 CGGAGAGCCTCGACAACATGAGCAAGCAGAGCTCATCAAGAGTACCTGGAACCTGAGAGA 886
Db 361 CGGAGAGCCTCGACAACATGAGCAAGCAGAGCTCATCAAGAGTACCTGGAACCTGAGAGA 420
OY 887 AGTGCCTCTCGGCGATGAGAGACGAAACACCGGCTGCGCTGAGAGCAAGCGGCTGG 946
Db 421 AGTGCCTCTCGGCGATGAGAGACGAAACACCGGCTGCGCTGAGAGCAAGCGGCTGG 480
OY 947 GTGGGAGCAGCGCGGTGTGGGAGCTGAGAGCTGAGAGCGGCTGCGCGCGGAGA 1006
Db 481 GTGGGAGCAGCGCGGTGTGGGAGCTGAGAGCTGAGAGCGGCTGCGCGCGGAGA 540
OY 1007 ACCTCCAGCTGCTGACC 1023
Db 541 ACCTCCAGCTGCTGACC 557

RESULT 15
LOCUS BE547281 745 bp mRNA linear EST 09-AUG-2000
DEFINITION 601073706F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3459960 5',
mRNA sequence.
ACCESSION BE547281
VERSION BE547281
KEYWORDS GI:9775926
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 745)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM8453 row: k column: 01
High quality sequence stop: 730.
Location/Qualifiers
1..745

FEATURES
Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3459960"
/clone_id="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.4 kb. Library prepared by Life Technologies.
BASE COUNT 195 a 168 c 276 g 106 t
ORIGIN
Query Match 49.9%; Score 539.2; DB 10; Length 745;
Best Local Similarity 93.6%; Pred. No. 5; 6e-100;
Matches 640; Conservative 0; Mismatches 33; Indels 11; Gaps 7;
OY 345 CGAGCTGCTCGCCCGCCCTTTGATGATCCGAGCGCAGTAAAGTTGGGGGCTCCTGCCG 404
Db 1 CGAGCTGCTCGCCCGCCCTTTGATGATCCGAGCGCAGTAAAGTTGGGGGCTCCTGCCG 58
OY 405 AGCGGCCCAAGAGAGAGTGGGCGACACACAGAGAGTGGGGGAAAGAAACATTAAGAG 464
Db 59 AAGGGGCCAAGAGAGAGTGGGCGACACACAGAGAGC-GGGGAAAGAAACATTAAGAG 117
OY 465 ACGCCCGTCAGAAAGAACCGGCATTTGAAACCGTACTACAAAGCTGAACTGGGAGAGAA 524
Db 118 ACGCCCGTCAGAAAGAACCGGCATTTGAAACCGTACTACAAAGCTGAACTGGGAGAGAA 176
OY 525 GAAAGAGTTCCAGAGAAACAGAGCCTTCGAGCTTCAAGAGATCCGAGCCGAGATTTGCC 584
Db 177 GAAAGAGTTCCAGAGAAACAGAGCCTTCGAGCTTCAAGAGATCCGAGCCGAGATTTGCC 236
OY 585 CAAGGCCAGCGCGGTGCGCCCTTAACACACAGCAGTTCCTCATGATGATCAACAGCA 644
Db 237 CAAGGCCAGCGCGGTGCGCCCTTAACACACAGCAGTTCCTCATGATGATCAACAGCA 296
OY 645 GGAGGAGCGGATCTCAAAACCGGCTTACTCCAAAGCGGGCGCCGCCCAATCCGAGCA 704
Db 297 GGAGGAGCGGATCTCAAAACCGGCTTACTCCAAAGCGGGCGG-CGCCAATCCGAGCA 355
OY 705 CACGAGCAGTATGACGATTCATGGAAGAGGGGGTATGAGAGATGGGGCAGCGATGGAT 764
Db 356 CACGAGCAGTATGACGATTCATGGAAGAGGGGGTATGAGAGATGGGGCAGCGATGGAT 415
OY 765 GGGAGGAGGAGCAGCAGGAGTTTCTGACGCGGAGCTTCGAGAGAGTACGAGCGGTACCA 824
Db 416 GGGAGGAGGAGCAGCAGGAGTTTCTGACGCGGAGCTTCGAGAGAGTACGAGCGGTACCA 475
OY 825 CACGAGAGCCTGCGAGAACATGAGCAAGCAGAGGCTCATCAAGAGTACCTGGAACCTGA 884
Db 476 CACGAGAGCCTGCGAGAACATGAGCAAGCAGAGGCTCATCAAGAGTACCTGGAACCTGA 535
OY 885 GAATGCTCTCTCGGCGATGAGAGAGAGAACACCGGCTGGCGCTGGAGACCAAGCGCT 944
Db 536 GAATGCTCTCTCGGCGATGAGAGAGAGAACACCGGCT--GGTGGAGAGCAAGCGGTT 592
OY 945 GGGTGGCGAGCAGCGCGGTGTGGCGGAGCTGGAGCTGGAGCCGCGTGGCGCGCGCA 1004
Db 593 GGGTG--GGAGGAGCGCGGTGTGGCGGAGCTGGAGCTGGAGTGGATTG-GGGCGGA 649
OY 1005 GAACCTCCAGCTGCTGACCGAGAA 1028
Db 650 GAACCTCCAGCTGCTGACCGAGAA 673

Search completed: June 18, 2003, 01:41:37
Job time : 1832 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 00:13:37 ; Search time 67 Seconds
(without alignments)
4943.446 Million cell updates/sec

Title: US-09-972-758A-1

Perfect score: 1080
Sequence: 1 atggccgagcattctgtc.....ttccaagtgtgagactag 1080

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	69	6.4	7218	1	US-08-232-463-14
2	68.4	6.3	1926	4	US-09-249-585A-2
3	68.4	6.3	2580	3	US-09-050-863-2
4	68.4	6.3	2580	4	US-09-359-081-2
5	68.4	6.3	5452	2	US-09-130-114-1
6	68.4	6.3	9600	4	US-08-910-647-1
7	68.4	6.3	9600	4	US-09-620-925-1
8	68.4	6.3	10596	1	US-07-884-811-15
9	68.4	6.3	10596	1	US-07-885-971-15
10	68.4	6.3	10596	1	US-08-087-783A-15
11	68.4	6.3	10596	2	US-08-194-088B-15
12	68.4	6.3	10596	2	US-08-194-087-15
13	68.4	6.3	10596	5	PCT-US93-04648-15
14	67.4	6.2	1926	4	US-09-130-114-2
15	67.4	6.2	1931	2	US-09-249-585A-4
16	65.8	6.1	9551	1	US-08-056-200-93
17	65.8	6.1	9551	2	US-08-800-644-93
18	63.4	5.9	3489	2	US-08-728-323A-1
19	63.4	5.9	3489	4	US-09-298-568-1
20	63.4	5.9	32207	2	US-08-770-379-20
21	63.4	5.9	32207	4	US-08-757-669A-20
22	63.4	5.9	32207	4	US-09-230-371A-20
23	60.2	5.6	2793	1	US-08-209-747-1
24	60.2	5.6	2793	1	US-08-458-298-1
25	58.4	5.4	2313	4	US-09-370-838-157
26	51.2	4.7	397	3	US-09-253-691-3
27	51	4.7	1995	1	US-08-425-069-3

28	51	4.7	1995	2	US-08-317-844B-3	Sequence 3, Appl
C 29	50.8	4.7	51259	3	US-08-781-891-209	Sequence 209, App
30	49.6	4.6	3138	4	US-09-434-408-1	Sequence 1, Appl
31	49	4.5	343	6	5273901-6	Patent No. 5273901
C 32	49	4.5	16442	3	US-08-781-891-208	Sequence 208, App
33	47.4	4.4	2277	1	US-08-676-967-5	Sequence 5, Appl
34	47.4	4.4	2277	1	US-08-676-974-5	Sequence 5, Appl
35	47.4	4.4	2277	2	US-09-098-487-5	Sequence 5, Appl
36	47.2	4.4	2338	1	US-08-425-069-1	Sequence 1, Appl
37	47.2	4.4	2338	2	US-08-317-844B-1	Sequence 1, Appl
38	46.8	4.3	234	1	US-08-469-802B-3	Sequence 1, Appl
39	46.8	4.3	234	2	US-08-267-803B-3	Sequence 3, Appl
40	46.6	4.3	1044	1	US-07-975-526-1	Sequence 1, Appl
41	46.6	4.3	1044	4	US-07-974-409C-422	Sequence 422, App
42	46.6	4.3	1044	4	US-07-974-409C-423	Sequence 423, App
43	46.2	4.3	1618	2	US-08-533-669A-9	Sequence 9, Appl
44	46.2	4.3	1618	2	US-08-607-509-1	Sequence 1, Appl
45	46.2	4.3	1618	2	US-08-454-036-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
City: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SFO ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9C-Fls
US-08-232-463-14
Query Match 6.4%: Score 69; DB 1; Length 7218;

Best Local Similarity 3.18; Pred. No. 1.7e-06;
Matches 12; Conservative 238; Mismatches 143; Indels 0; Gaps 0;

[illegible]

```

RESULT 2
US-09-249-585A-2
: Sequence 2, Application US/09249585A
: Patent No. 6417002
: GENERAL INFORMATION:
: APPLICANT: Horlick, Robert
: TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISODES
: FILE REFERENCE: 0867/OD905
: CURRENT APPLICATION NUMBER: US/09/249,585A
: CURRENT FILING DATE: 1999-02-11
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 2
: LENGTH: 1926
: TYPE: DNA
: ORGANISM: Epstein Barr Virus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1926)
: OTHER INFORMATION: coding strand of EBNA-1 DNA
: US-09-249-585A-2

```

Query Match	6.38;	Score 68.4;	DB 4;	Length 1926;
Best Local Similarity	44.08;	Pred. No. 1.7e-06;		
Matches 288; Conservative	0;	Mismatches 366;	Indels 0;	Gaps 0;

[illegible]

QY	646	AAGGAGCCGGATCTCATAAACCGGCTTATCTATCTCAAGGGGGGCGGCGCCAAATTCGAGAC	705
Db	545	GAGGGGCGAGAGCAGAGAGGGGCGAGGGGGCGAGGGGCGAGCAGAGAGGAGGGG	604
QY	706	ACCAGCAGTCAGCACTTCATGGAAGAAAGGGGGTGAGAGATGGGGGCCAGCATGGGATG	765
Db	605	CAGAGAGCAGAGGGGCGAGAGGGGCGAGGGGCGAGAGCAGAGAGGGCGAGGAGGAGGAG	664
QY	766	GGAGGGGACGCGCAGGAGTTTCTGACGCGGACCTTCTGGAGACGTACGAGCGGTACCAC	825
Db	665	GAGGGGCGAGGAGGGGCGAGGGGCGAGGACGAGAGGGGCGAGGAGCAGAGAGGGGCGAGGAG	724
QY	826	ACGGAGACCCCTGCAAGAACATGACGACAGCAGGCTCATCAAGGAGTACTTACTGAACTGGAG	885
Db	725	CAGGAGGGGCGAGAGCAGAGAGGGGCGAGGGGCGAGGACAGAGGGGCGAGAGGGGCGAG	784
QY	886	AAGTATCTCTTGCCATATGAGGAGCAGCAAAACCGGCTGCGGGCTGGAAGAGCAAGCGCTG	945
Db	785	GAGCAGAGAGGGGCGAGAGGGGCGAGGACGAGGAGGAGGGGCGAGAGGGGCGAGAGGAGGAG	844
QY	946	GGTGGCGACGACGCGCGGTGTGCGGGGAGCTGAGGCTGAGGCTGAGACCGGCTGCGCGCGAG	1005
Db	845	GAGGGGCGAGGAGGGGCGAGAGCAGAGGCGGCGAGGAGGCGAGGAGCAAGAGGGGCGAGGAG	904
QY	1006	AACTTCACGTCTGACCGGAGAAACGAATTGACCGGCGAGCGAGGAGCGAGCGCCG	1059
Db	905	GGGCGAGGCGAGAGGGGCGAGAGGGGCGAGGAGCAGAGGAGAGGGGCGAGGAGCG	958

RESULT 3
 US-09-050-863-2
 : Sequence 2, Application US/09050863
 : Patent No. 614111
 : GENERAL INFORMATION:
 : APPLICANT: Lao, Ying
 : APPLICANT: Hiang, Betty
 : APPLICANT: Payan, Don
 : TITLE OF INVENTION: Mammalian Protein Interaction Cloning
 : TITLE OF INVENTION: System
 : NUMBER OF SEQUENCES: 5
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSSEE: Flehr, Hohbach, Test, Albritton & Herbert
 : STREET: 4 Embarcadero Center, Suite 3400
 : CITY: San Francisco
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94111-4187
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/050,863
 : FILING DATE: 30-MAR-1998
 : CLASSIFICATION:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Silva, Robin M.
 : REGISTRATION NUMBER: 38,304
 : REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 781-1989
 : TELEFAX: (415) 949-8711
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2580 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: unknown
 : TOPOLOGY: unknown
 : MOLECULE TYPE: DNA
 : US-09-050-863-2

US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Hotlick, Robert A.
; APPLICANT: Dama, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130.114
; CURRENT FILING DATE: 1998-08-06
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: YEBA
US-09-130-114-1

Query Match 6.3%; Score 68.4; DB 2; Length 5452;
Best Local Similarity 44.0%; Pred. No. 2.1e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGCGAAGAGAGTGGGACAGCAGACAGCTGGGGAAGAAAACTAAGAGA 465
DB 2117 GGGCAGAGAGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2058
QY 466 CGCCCGTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525
DB 2057 GAGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1998
QY 526 AAAAGTTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585
DB 1997 CAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1938
QY 586 AAGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 645
DB 1937 GGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1878
QY 646 GAGAGCCGAGATCTCAAAACCGGCTGTACTTCAAGAGAGAGAGAGAGAG 705
DB 1877 GAGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1818
QY 706 ACCAGCGATGACACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
DB 1817 CAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1758
QY 766 GAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825
DB 1757 GAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1698
QY 826 ACGAGAGCGCTGTGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
DB 1697 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
QY 886 AAGTGCCTCTCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 945
DB 1637 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
QY 946 GGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005
DB 1577 GAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
QY 1006 AACCTCAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1059
DB 1517 GGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464

RESULT 6
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433

GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 923-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-910-647-1

Query Match 6.3%; Score 68.4; DB 4; Length 9600;
Best Local Similarity 44.0%; Pred. No. 2.4e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGCGAAGAGAGTGGGACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465
DB 734 GGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 793
QY 466 CGCCCGTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525
DB 794 GAGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 853
QY 526 AAAAGTTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585
DB 854 CAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 913
QY 586 AAGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 645
DB 914 GGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 973
QY 646 GAGAGCCGAGATCTCAAAACCGGCTGTACTTCAAGAGAGAGAGAGAGAGAG 705
DB 974 GAGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1033
QY 706 ACCAGCGATGACACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
DB 1034 CAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1093
QY 766 GAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825
DB 1094 GAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1153
QY 826 ACGAGAGCGCTGTGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
DB 1154 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1213

QY	886	AAGGCTCTTCGGCCATGGAGAGACAGAAACACGGGCTGGGGCTGGAGACAGCAAGGGGCTG	945
Db	1214	GAGCAGGAGGGGGCCAGGAGGGGCCAGGACCGAGGAGGAGGGGCCAGAGGGCCAGAGGACGAGAG	1273
QY	946	GGTGGCCACGACGCGCGCTGTGCGGGAGCTGGAGCTGGAGCTGGACCGGCTCGGCCCGCAG	1005
Db	1274	GAGGGGCGACGAGGGGGCCAGGACCGAGAGGGGCGAGAGGGGCCAGGACAGGAGGGGCGCAGGAG	1333
QY	1006	AACCTCCAGCTGCTGACCGAGAACGAACTGTCACGGCGACGAGAGCGAGCGCGC	1059
Db	1334	GGGCGAGAGGACGAGGAGGGGCGAGGAGGGCGCAGGACGAGGAGGAGGGGCCAGGACGAG	1387

RESULT 7
US-09-62

Sequence 1, Application US/09620925
Patent No. 6468986
GENERAL INFORMATION:
APPLICANT: Zuckermann et al.
TITLE OF INVENTION: Compositions and Methods for Polynucleotide Delivery
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/620,925
FILING DATE: 21-Jul-2000
CLASSIFICATION: <Unknown>
Prior Application DATA:
APPLICATION NUMBER: 08/910,647
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fujita, Sharon M.,
REGISTRATION NUMBER: 38,459
REFERENCE/DOCKET NUMBER: 1218.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-2706
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
DS-09-620-925-1

Query Match	6.38;	Score 68.4;	DB 4;	Length 9600;
Best Local Similarity	44.08;	Pred. No. 2.4e-06;		
Matches 288; Conservative	0;	Mismatches 366;	Indels 0;	Gaps 0;

QY	400	GGGGGCGAAGACAGAGCTGGGGGACAGCAGCAGCAGAGACAGCTGGGGAAGAAAAACATTAAGAGA	465
Db	734	GGGACAGACGACGAGAGGGCGCCAGGAGCAGGAGAGCGGGCAGGAGAGGGGCGACAGAGGGGCGAG	793
QY	466	CGCCCTCCAAACAGAGCGCGATTGGAAACCGTACTCAAGCTGAACTGGGAAGACAAAG	525
Db	794	GAGGGCAGAGACGAGAGGAGGGCGAGGAGCGCAGGAGGAGGGGCGAGAGGGGCGACGAGAGGGG	853
QY	526	AAAAAGTTGACGAGAAACAGAGCCCTTGAGCTTCAAGGATCCGAGCCGAGATTTTGGCC	585
Db	854	CAGCAGCAGAGAGAGGGGCGAGGACGAGAGGAGGAGGCGCAGGAGGGGCGAGACGAGAGGAG	913

OY	586	AAGGGCCACGCGGCTGGGCGCCCTATTAACACCCAGCGAGTTTCTCATGGATGATCAACGAC	645
Db	914	GGCGACGAGAGGGGCGACGAGAGGGCGACGAGCAGAGAGGGGCGAGAGCAAGAGAGGGGCGAC	973
OY	646	GAGGAGCCGGATCTCAAAACCGCGCTTACTCCAAAGCGGGCGCGCCCAATTCGACGAC	705
Db	974	GAGGGGCGAGGAGCGAGGAGAGGGCGAGAGGGGCGAGAGCGGCGACGACAGAGAGAGGGG	1033
OY	706	ACCAAGCATACGACTTCATCGAAGAAAGGGGTTCAGCAGATGTGGGGCGACGATGGGATG	765
Db	1034	CAGGAGCAGGAGGGGCGACGAGGGGCGAGAGGGGCGAGAGCGAGGAGGGGCGACGACGAGAG	1093
OY	766	GGAGGGCAGCGACGACGAGTTTCTTCACCGGAGTTCCTGAGAGAGTACGAGCGGTACAC	825
Db	1094	GAGGGGCGAGGAGGGGCGACGAGAGGGCGAGGAGCGAGAGGGGCGAGAGCAAGAGAGGGCGACGAG	1153
OY	826	ACCGAGAGCCTTCAGAAACATGACCAAGCAGAGACTCATCAAGAGATACCTTGGAACTGGAC	885
Db	1154	CAGGAGGGGCGACGAGAGCGAGGAGGGGCGAGAGGGGCGAGAGCGACGAGAGGGGCGACGAGGGGCGACG	1213
OY	886	AAGTGGCCTTCGCGCATGAGAGAGGAGAGCAACCGGCTGCGCGCTGCGAGAGCAACGCGGCTG	945
Db	1214	GAGCAGGAGGGGCGACGAGAGGGGCGACGAGCAGGAGAGAGGGGCGAGAGGGGCGACGAGAGGAG	1273
OY	946	GGTGGCAGCAGACCGCGCTGTCCGGAGCTGAGCTGAGCTGAGACCGGCTGGCGCCGAG	1005
Db	1274	GAGGGGCGAGGAGGGGCGAGAGCGAGAGAGGGGCGAGAGGGGCGAGAGCGACGAGAGCGACGAGAG	1333
OY	1006	AACCTTCAGCTGTGACACCGAGAAACAGAACCTGACCGGAGCGAGCGAGCGACGCGCGC	1059
Db	1334	GGGTCAGACGACGAGAGGGGCGAGAGGGGCGAGAGCGAGGAGGAGGAGGAGGAGGAGGAGCAG	1387

RESULT 8

US-07-884-811-15
Sequence 15, Application US/07884811
Patent No. 5316921
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Matk, Melanie R
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patla (genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/884, 811
FILING DATE: 19920518
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dregger, Ginger R.
REGISTRATION NUMBER: 33, 055
REFERENCE/DOCKET NUMBER: 755.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/223-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ. ID NO.: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1036 bases
TYPE: NUCLEIC ACID

Db	2328	GAGGGGACGAGCAGAGGAGGCGGACGAGCAGGAGGGGCAGAGGGGCCAGGAGGGG	2387
Oy	526	AAAAAGTTCCAGCAGAAACAGACCTTCGAGCTTCAAGATCCGAGACCAGATGTTGCC	585
Db	2388	CAGAGAGCAGGAGGAGGGGCAGGACAGGAGAGGGGCGAGGGGCGACAGACAGAGAG	2447
Oy	586	AAGGCGCCGCGGTGCGCCCTATTACACACAGCATGTTCTCATGGATGATCACAGCAG	645
Db	2448	GGCGACGAGAGGGGCGCAGGAGGGGCGAGAGCAGAGAGGGGCGAGAGCAGAGGGGCGAG	2507
Oy	646	GAGAGGCCGGAATTCATAAACCCGCTGTACTCCAAGCGGGCCGCCCAAATCCGACAGC	705
Db	2508	GAGGGGCGAGGAGCAGGAGGAGGGCGACAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGG	2567
Oy	706	ACCAAGCATGACGACTTCATGGAAGAAGGGGGTGAGAGGATGAGGGGCGAGGATGCGATG	765
Db	2568	CAGAGACGAGAGGGGCGCAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGCGACAGAG	2627
Oy	766	GGAGGGGACGCGACGAGTTCCTTCAGCGGGGACTTCTCGAGAGCTACGACGCGTACCAC	825
Db	2628	GAGGGGCGAGAGGGGCGACGAGGGGCGAGAGGCGAGAGAGGGGCGAGAGGAGGCGAGAG	2687
Oy	826	ACGAGAGACCTTCGACAACATACGACAGCAGAGACTCTATCAAGAGATACCTTGGAACTGAG	885
Db	2688	CAGAGAGGGGCGAGGACGAGGGGCGAGAGGGGCGAGGAGCAGAGAGGGGCGAGAGGGCGAG	2747
Oy	886	AAGTGCCTCTCGCCCATGAGAGCAGAGCAACCGGCTGCGGGCTGGAAGCAGAGCGGCTG	945
Db	2748	GAGCAGAGAGGGGCGCAGAGAGGGGCGAGAGGAGAGGAGAGGGGCGAGAGGGGCGAGAGGAG	2807
Oy	946	GGTGGCCACGACGCGCGGTGTGCGGGAGCTGGAAGCTGGAAGCTGAGCAGGCTCGCGCGAG	1005
Db	2808	GAGGGGCGAGGAGGGGCGACGAGACAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAG	2867
Oy	1006	AACMTCGACGTGTGACCGCAGACAAGAACTGCACCGGCGAGCGAGCGAGCGCCG	1059
Db	2868	GGCGAGGAGCAGAGAGGGGCGCAGAGGGGCGTAGAGCAGAGGAGGGGCGAGGAGCAG	2921
 RESULT 12 US-08-194-087-15			
Sequence 15, Application US/08194087			
Patent No. 5879910			
GENERAL INFORMATION:			
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.			
TITLE OF INVENTION: HEPATOCTYCE GROWTH FACTOR PROTEASE DOMAIN VARIANTS			
NUMBER OF SEQUENCES: 21			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Genentech, Inc.			
STREET: 460 Point San Bruno Blvd			
CITY: South San Francisco			
STATE: California			
COUNTRY: USA			
ZIP: 94080			
COMPUTER READABLE FORM:			
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: patin (Genentech)			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/194,087			
FILING DATE: 18-MAY-1992			
CLASSIFICATION: 435			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER:			
FILING DATE:			
ATTORNEY/AGENT INFORMATION:			
NAME: Dreger, Ginger R.			
REGISTRATION NUMBER: 33,055			
REFERENCE/DOCKET NUMBER: 779			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: 415/225-3216			

```

: TELEFAX: 415/952-9881
: TELER: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 10596 bases
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: US-08-194-087-15

Query Match      6.3%  Score 68.4; DB 2; Length 10596;
Best Local Similarity 44.0%; Pred. No. 2.5e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

OY 406 GGGGGCCGAAGAGAGTGGCGACACAGACAGACAGCTGGGGAAGAAAAACATAGAGA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2268 GGGCAGGAGACGACAGAGAGAGAGGGCCAGGACGAGAGAGGGCCAGAGAGGGCCAG 2327
OY 466 CGCCCTTCACAGAAAGAAAGCGCATTTGGAACCGTACTCAAGCTGAACCTGGGAAGAG 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2328 GAGGGGACAGAGCGACGAGAGAGGCGCAGAGACGAGAGAGGGCCAGAGAGGGCCAG 2387
OY 526 AAAAAAGTTTCGACGAGAAACAGAGCCTTCGACCTTCAGAGATCCGAGCCAGATTCGCC 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2388 CAGAGACAGAGGGAGGGGCGACAGACAGAGAGAGGGGCGACAGAGGGGCGACAGAGAG 2447
OY 586 AAGGGCCAGCCGCTGCGCCCTTATACACACGACGATTCCTCATGATGATCAGACACG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2448 GGGCAGAGAGGGGCGACAGAGGGCGCAGAGACAGAGAGGGGCGACAGAGAGAGGGGCG 2507
OY 646 GAGAGACCGGATCTCCAAAACCGGCTGTACTTCGAAAGCGGGCGCCGCCCAATCCGAGC 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2508 GAGGGGCGACGAGCGACGAGAGAGGGGCGACAGGGGCGACAGAGGGGCGACAGAGAG 2567
OY 706 ACCAGCGATGACGACTTCATGAAAGAGGGGGGTAGAGGATGGGGCGACGATGGGATG 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2568 CAGAGACAGAGAGGGGCGACAGAGGGGCGACAGAGGGGCGACAGAGAGGGGCGACAGAG 2627
OY 766 GAGAGGGGACGGGCGACGAGTTTCTGACGCGGACATTCTCGGAGACGTACGACGCTACC 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2628 GAGGGGCGACGAGGGGCGACGAGGGGCGACGAGAGGGGCGACGAGAGGGGCGACAGAG 2687
OY 826 ACGGAGAGCCCTCGAAGAACATGACGACGAGGAGCTCATCAAGGAGTACCTGACACTGG 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2688 CAGAGAGGGGCGACGAGACGAGAGGGGCGACAGAGGGGCGACGAGAGGGGCGACAGAG 2747
OY 886 AAGTGCCTCTTCGCGCATGAGAGACGAGAAACCGGCTCGGCTGGAGAGCAAGCGGCTG 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2748 GAGCAGGAGAGGGGCGACGAGAGGGGCGACGAGAGGGGCGACGAGAGGGGCGACAGAG 2807
OY 946 GGTGGCCACGACGCGGCTGTGCGGGGAGGTGGAGCTGAGACTGAGACCGGCTCGGCGCGAG 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2808 GAGGGGCGACGAGGGGCGACGAGAGGGGCGACGAGAGGGGCGACGAGAGGGGCGACAGAG 2867
OY 1006 AACCTCCAGCTGTGACCGAGAACGAACTGACCGGCGACGAGAGGAGCGAGCGCG 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2868 GGGCAGGAGCAGAGAGGGGCGACGAGAGGGGCGACGAGAGGGGCGACGAGAGCAG 2921

RESULT 13
PCT-US93-04648-15
: Sequence 15, Application PC/TUS9304648
: GENERAL INFORMATION:
: APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melanhi
: TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080

```


